UNDERSTANDING HTLV-I ENZYMOLOGY
&
PREPARATION AND CHARACTERIZATION OF LEAD INHIBITORS FOR THE TREATMENT OF HTLV-I INFECTION

A Dissertation
Presented to
The Academic Faculty

By

Kelly Joy Dennison

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Chemistry

Georgia Institute of Technology

December 2005

UNDERSTANDING HTLV-I ENZYMOLOGY
&
PREPARATION AND CHARACTERIZATION OF LEAD INHIBITORS FOR THE TREATMENT OF HTLV-I INFECTION

Approved by:

Dr. Suzanne B. Shuker, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Thomas M. Orlando, Co-Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Donald F. Doyle
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. C. David Sherrill
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Andrew S. Bommarius
School of Chemical and Biochemical Engineering
Georgia Institute of Technology

Dr. S. Michele Owen
National Center for HIV, STD, and TB Prevention
Centers for Disease Control & Prevention

Dr. Vicky L. H. Bevilacqua
Edgewood Chemical Biological Center
US Army

Date approved: August 11, 2005

ACKNOWLEDGEMENTS

I would like to thank the many people that supported me through this process. To all my family: my parents, Chuck and Joy; my sisters, Gerri and Laurel; and my brother Keary, as well as Kent, Brian, and Brandy; the Chapels; the Bonuras; Tekkie and Chet; Auntie Marie; and my nephews and niece, Patrick, Ryan, Kevin, Kaitlyn, and Riley. Thanks for all your encouragement and/or financial support and most of all, for believing in me. To my friends: thanks for being present. To all the Shuker Group members: Dr. Suzanne B. Shuker, Dr. David Gaul, Dr. Kevin Caran, Dr. Julie Ha, Dr. Vicki Mariani, Dr. Bryan Herger, Dr. Beth Brewster, Dr. Josh Sasine, Dr. Ta-Tanisha Favor, Micah Gliedt, David Kim, Suazette Reid, Bob Po Chen, and soon-to-be Dr. Mike Kulis, thanks for all your input. To all those in Wing 3A and IB^2: it was great getting to work with you. To undergraduate student Alana Moore: I hope all your dreams come true and thanks for inspiring me as much as I inspired you.

Finally, I would like to dedicate this work to those that have passed on before me: Grammy, Uncle Rodney, Uncle Butch, and Luz Hamilton. May peace be upon you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartic acid
ATL	Adult T-cell leukemia/lymphoma
ATLV	Adult T-cell leukemia virus
BLV	Bovine leukemia virus
CA	Capsid
Cys, C	Cysteine
Dabcyl	4-(4-Dimethylaminophenylazo)benzoyl
DCC	N,N'-Dicyclohexylcarbodiamiide
DIPEA	Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
EDANS	5-(2-aminoethylamino)-1-napthalenesulfonic acid
FIV	Feline immunodeficiency virus
Fmoc	9-Fluorenylmethoxycarbonyl
Gln, Q	Glutamine
Glu, E	Glutamic acid
Gly, G	Glycine
HBTU	2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
His, H	Histidine

HIV-1	Human immunodeficiency virus type 1
HMBA	4-hydroxymethylbenzoic acid
HOBt	N-hydroxybenzotriazole
HTLV-I	Human T-cell leukemia virus type I
Iaa	Isoamylamine
IN	Integrase
Ile, I	Isoleucine
Iva	Isovaleryl
Leu, L	Leucine
Lys, K	Lysine
MA	Matrix
Met, M	Methionine
MOE	Molecular Operating Environment
Nph	4-Nitrophenylalanine
NC	Nucleocapsid
ORF	Open reading frame
PBMC	Peripheral mononuclear cells
PCR	Polymerase chain reaction
PDB	Protein data bank
Phe, F	Phenylalinine
PR	Protease
Pro, P	Proline
RSV	Rous sarcoma virus
RT	Reverse transcriptase

Ser, S	Serine
Sta	Statine
TEA	Triethylamine
TBTU	2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate
Thr, T	Threonine
TPCK	N-Tosyl-L-phenylalanine chloromethyl ketone
Trp, W	Tryptophan
TSP/HAM	Tropical spastic parapesis/HTLV-I associated myelopathy
Tyr, Y	Tyrosine
Val, V	Valine

SUMMARY

The primary goals of our research are to understand the virology and enzymology of human T-cell leukemia virus type I (HTLV-I) that will lead to the development of treatments for patients infected with HTLV-I. HTLV-I is an oncogenic virus of the *Retroviridae* family and is the causative agent of adult T-cell leukemia/lymphoma (ATL), tropical spastic paraparesis/HTLV-I associated myelopathy (TSP/HAM). HTLV-I has been classified as a dangerous emerging pathogen by the Centers for Disease Control & Prevention with at least 20 million people infected with the virus. This is a significant problem because there are currently no effective treatments to control HTLV-I infection and prevent or treat HTLV-I induced ATL and TSP/HAM.

The protease is necessary for retroviral maturation and replication and is, therefore, an attractive target for inhibitor design. Investigation of peptide mimetic compounds incorporating the tetrahedral intermediate of aspartyl protease catalyzed cleavage are crucial for the development of lead inhibitors. Compounds containing statine, 4-amino-3-hydroxy-5-phenylpentanoic acid (AHPPA), or hydroxyethylamine (HEA) are presented in this work. The best compound was a statine-based inhibitor, which had a K_i = 29 ± 4 nM and 88% inhibition against an HTLV-I protease native substrate in a FRET assay.

CHAPTER 1

HUMAN T-CELL LEUKEMIA VIRUS TYPE I

1.1 Human T-Cell Leukemia Virus Type I

Human T-Cell Leukemia Virus Type I (HTLV-I) is an oncogenic virus of the *Retroviridae* family (Murphy & Kingsbury, 1990). HTLV-I has been etiologically implicated as the causative agent of adult T-cell leukemia/lymphoma (ATL) (Gitlin et al., 1993), tropical spastic paraparesis/HTLV-I associated myelopathy (TSP/HAM) (Gessain et al., 1985), and other diseases such as uveitis, infectious dermatitis, polymyositis, and polyarthritis (Gessain, 1996).

In 1978, HTLV was isolated from cultured $CD4^+$ T-lymphocytes from a patient in the United States with cutaneous T-cell lymphoma (Gallo & Thomson, 1996; Poiesz et al., 1980). A second isolate was obtained from a different patient with $CD4^+$ T-cell leukemia (Gallo & Thomson, 1996). Independently, a similar virus described as Adult T-Cell Leukemia Virus (ATLV) was isolated from an adult T-cell leukemia/lymphoma patient in Japan. The name HTLV-I was adopted for both HTLV and ATLV as a result of the homology of the viral genome and antigens (Gallo & Thomson, 1996; Gitlin et al., 1993; Yoshida et al., 1982; Yoshida et al., 1984).

HTLV-I is endemic in southwestern Japan (Gallo & Thomson, 1996; Green & Chen, 1993), as well as areas in the Carribean, sub-Saharan Africa, Seychelles Islands, Melanesia, Australia, Philippines, the Americas, south India, and southwestern Asia (Gallo & Thomson, 1996). Worldwide estimates are that approximately 20 million people are infected with HTLV-I (Dekaban et al., 1992). In the United States, there is an estimated prevalence of 0.016-0.1% HTLV-I infection among blood donors (Zucker-Franklin & Pancake, 1998). Transmission is vertical via maternal breast milk, intravenous drug use with contaminated sharps, transfusion with infected blood, and sexual transmission (Gessain, 1996; Kaplan & Khabbaz, 1993; Khabbaz et al., 1990). Sexual transmission is much more efficient from male to female (60.8%) than from female to male (0.4%) (Gitlin et al., 1993), therefore, a disproportionately higher number of females are infected with HTLV-I (Gessain, 1996).

1.2 Transmission of HTLV-I Virions

The model of retroviral infection suggests that virus particles recognize and bind to specific target cells. This recognition and binding occurs through interactions between the envelope glycoprotein complex (Env) and a surface receptor. This results in adhesion of the cell and the viral membranes, which leads to the virus core being deposited into the host cell cytoplasm (Derse & Heidecker, 2003). This model has recently been challenged for HTLV-I. It appears that cell-to-cell contact is essential for viral transmission (Derse & Heidecker, 2003; Igakura et al., 2003).

The viral envelope glycoprotein complex consists of two proteins, the glycosylated hydrophilic surface protein (SU) and a transmembrane protein (TM). Together, these proteins form the spikes on the surface of the virus. Both proteins are proteolytically cleaved from the Env precursor protein upon transfer to the cell surface. Neither of these proteins are necessary for envelope assembly, however they do have a role in the entry process (Coffin et al., 1997). The SU specifically binds to a specific cell receptor. This binding association appears to trigger the membrane fusion-inducing potential of the TM, which then results in the fusion of cell and virus membranes (Derse & Heidecker, 2003; Coffin et al., 1997).

The precise mechanism of cell-to-cell infection is not well understood. The receptor to which the Env bind has not been identified. However, it has been demonstrated that the HTLV-I core (Gag) proteins along with the HTLV-I genome accumulate at the cell-to-cell junction and are subsequently transferred to an uninfected lymphocyte (Igakura et al., 2003).

1.3 Pathogenesis

To date, HTLV-I is the only retroviral agent known to cause human cancer. HTLV-II, which is the other known human oncovirus, is not linked with cancer. Like HIV, HTLV-I targets $CD4^+$ T-cells. The main pathological difference between HTLV-I and HIV is that HTLV-I alters the cells, which may result in tumor growth and other neurological disorders, while HIV destroys $CD4^+$ T-cells. Both HTLV-I and HIV have regions in their genomes that encode Gag, Pol, and Env polyprecursors, as well as accessory proteins that appear to be necessary for viral infectivity *in vivo*. The 3' end of HTLV-I has a *pX* region that has four open reading frames (ORFs), (Johnson and Franchini, 2002). It is this *pX* region that has placed HTLV-I into a unique class of retroviruses that includes HTLV-II and bovine leukemia virus (BLV), (Yoshida and Seiki, 1990), which is the etiologic cause of enzootic bovine leucosis (Coffin *et al.*, 1997). The first ORF codes for a small hydrophobic protein, $p12^I$. This protein, similar to HIV-1 Nef, is membrane-associated and is required for *in vivo* infectivity but not *in vitro* replication (Johnson and Franchini, 2002).

HTLV-I has the ability to infect a wide range of cell types. However, it is the T-cells and B-cells that are prone to transformation. Therefore, the T-cell lymphocyte tropism appears to be induced by post-infectivity events rather than at the receptor level (Coffin *et al.*, 1997).

1.4 Diseases Associated with HTLV-I

ATL is a malignancy of mature $CD4^+$ T-lymphocytes and is characterized by "lymphoma/leukemia often associated with generalized lymphadenopathy, hypercalcemia, infiltrative skin lesions, hepatosplenomegaly, bone marrow involvement, and lytic bone lesions of the skull and long bones" (Gitlin *et al.,* 1993). Malignant lymphocytes found in the peripheral blood of HTLV-I infected individuals can be characterized by irregular nuclei referred to as "flower cells." ATL is a progressive disease with four stages: pre-ATL (preleukemia), smoldering ATL, chronic ATL, and acute or subacute ATL. Pre-ATL individuals are typically diagnosed with abnormal lymphocytes. However, few apparent symptoms are displayed. An estimated 30% of pre-ATL patients progress to the smoldering and chronic ATL stages with indicative skin lesion formation. Further progression to the acute stage may occur in a few months and is indicated by aggressive $CD4^+$ T-lymphocyte malignancy (Gitlin et al., 1993). ATL is the most prevalent disease caused by HTLV-I with 20-30 years for progression following viral infection (Green & Chen, 1993).

TSP/HAM is a neurological demyelinating disease. Affected individuals generally display abnormal muscle movement of the extremities and some sensory loss (Gitlin et al., 1993). TSP/HAM differs from ATL in that individuals may develop TSP/HAM disease in a matter of a few years following infection with HTLV-I rather than 20-30 years of the ATL progression (Green & Chen, 1993).

1.5 Current Treatment of HTLV-I

Currently, there is no adequate treatment for HTLV-I infected individuals. Combination therapies have been tried, with limited success. For example, the OPEC/MPEC treatment protocol (weekly doses of vincristine, methotrexate, prednisolone, etoposide, and cyclophosphamide) or daily doses of etoposide and prednisolone (DOEP) has had a limited impact. The median survival time (MST) of all infected patients and those with acute leukemia, lymphoma and progressive chronic leukemia was 7.5-12.4 months. The MST for patients treated with the OPEC/MPEC regimen was 7.1 months while the MST for patients taking the DEOP regimen was 18.0 months (Matsushita et al., 1999).

1.6 Significance

HTLV-I was the first human oncogenic retrovirus to be discovered. Even though HTLV-I has been classified as a dangerous emerging pathogen by the Centers for Disease Control & Prevention, it is not as well understood as HIV-1 (Ewald, 1996; Satcher, 1995). Approximately 10% of the estimated 15-20 million individuals infected with HTLV-I will develop the fatal disease ATL. (Gessain, 1996; Zucker-Franklin & Pancake, 1998). This is a significant problem because there are currently no effective

treatments to control HTLV-I infection and prevent or treat HTLV-I induced ATL and TSP/HAM.

The protease is necessary for retroviral maturation and replication and is, therefore, an attractive target for inhibitor design. However, not much is known about the structure and enzymology of HTLV-I protease. It is crucial for the designing of inhibitors of HTLV-I protease to have a structural understanding of the sterics and electronic interactions necessary for binding to HTLV-I protease active site in order to design high affinity inhibitor drugs. Since there is no known structure of HTVL-I protease, a theoretical model of this protease was used to better understand the binding interactions of the active site and thus facilitate design of effective inhibitors (Dennison et al., 2003; Herger, 2004).

Investigation of peptide mimetic compounds that incorporate the tetrahedral intermediate of aspartyl protease catalyzed-cleavage is can provide for the development of lead inhibitors. Compounds containing statine, 4-amino-3-hydroxy-5-phenylpentanoic acid (AHPPA), or hydroxyethylamine (HEA) are expected to be potent protease inhibitors with $K_i < 1$ nM.

HTLV-I protease contains ten C-terminal residues that are not conserved across non-leukemic retroviral proteases. Leu^{115} and Pro^{116} form a prototypical cleavage site (L/P) in the C-terminal tail domain and should result in autoprocessing comparable to the non-leukemic proteases. The cleavage at this site is inhibited by Tyr^{114} or Glu^{117}, which occupy the putative P2 and P2' substrate position, respectively. Catalysis experiments with HTLV-I and peptide mutants incorporating mutations in the P2 and P2' (Tyr and Glu, respectively) could provide answers to why this C-terminal tail remains intact and could provide additional understanding of HTLV-I protease enzymology.

A second site for treatment is the HTLV-I nucleocapsid protein. Not much is known about the structure of the nucleocapsid protein that is processed from the Gag precursor polyprotein. The primary role of the nucleocapsid is in recognition and packaging of the viral genome (Lee et al., 1998; Amarasinghe et al., 2000; Williams et al., 2002; Coffin et al., 1997). HIV-1 nucleocapsid structure has been determined and several nucleocapsid inhibitors are being developed (Amarasinghe et al., 2000; Lee et al., 1998; Stephen et al., 2002). However, the structure of HTLV-I nucleocapsid has not yet been reported. Determining the three-dimensional structure of this important protein would enable the development of alternative drug therapies for the treatment of HTLV-I infections.

The primary aims of this work were to examine the protease activity of synthesized 10-mer peptides that incorporate a possible C-terminal tail cleavage site, to explore the *in vivo* behavior of HTLV-I L40I mutant protease and the L40I mutant without the last 10 C-terminal residues (L40IΔ10); to prepare pentapeptides incorporating statine, AHPPA, or HEA isosteric mimetics of the tetrahedral intermediate of aspartyl protease-catalyzed peptide cleavage; to test inhibition of the peptide mimetics in an *in vitro* cell-culture assay; and to determine the structure of HTLV-I nucleocapsid.

REFERENCES

Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R. & Summers, M. F. 2000. NMR Structure of the HIV-1 Nucleocapsid Protein Bound to Stem-Loop SL2 of the ψ-RNA Packaging Signal. Implications for Genome Recognition. *J.M.B.*, **301,** 491-511.

Coffin, J. M., Hughes, S. H. & Varmus, H. E. 1997. *Retroviruses*. Plainview: Cold Spring Harbor Laboratory Press.

Dekaban, G. A., King, E. E., Waters, D. & Rice, G. P. 1992. Nucleotide sequence analysis of an HTLV-I isolate from a Chilean patient with HAM/TSP. *AIDS Res Hum Retroviruses*, **8,** 1201-7.

Dennison, K. J., Herger, B. E. & Shuker, S. B. 2003. Protein Data Bank ID 1O0J.

Derse, D. & Heidecker, G. 2003. Forced entry--or does HTLV-I have the key? *Science*, **299,** 1670-1671.

Ewald, P. W. 1996. Guarding against the most dangerous emerging pathogens: Insights from evolutionary biology. *Emerg. Infect. Dis.*, **2,** 245-256.

Gallo, R. C. & Thomson, M. M. 1996. Introduction. In: *Human T-Cell Lymphotropic Virus Type I* (Ed. by Hollsberg, P. & Hafler, D. A.). Chichester, UK: Wiley.

Gessain, A. 1996. Epidemiology of HTLV-I and associated diseases. In: *Human T-Cell Lymphotropic Virus Type I* (Ed. by Hollsberg, P. & Hafler, D. A.), pp. 31-64. Chichester, UK: Wiley.

Gessain, A., Vernant, J. C., Maurs, L., Barin, F., Gout, O., Calender, A. & De The', G. 1985. Antibodies to human T-cell leukemia virus type I in patients with tropical spastic parapesis. *The Lancet*, **2,** 407-410.

Gitlin, S. D., Dittmer, J., Reid, R. L. & Brady, J. N. 1993. Human Retroviruses. (Ed. by Cullen, B. R.). New York: Oxford University Press.

Green, P. L. & Chen, I. S. Y. 1993. The Retroviridae. (Ed. by Levy, J. A.). New York: Plenum Press.

Herger, B. E., Mariani, V. M., Dennison, K. J., Shuker, S. B. 2004. The Ten C-Terminal Residues of HTLV-I Protease Are Not Necessary for Enzymatic Activity. *Biochemical and Biophysical Research Communications*, **In press**.

Igakura, T., Stinchcombe, J. C., Goon, P. K. C., Taylor, G. P., Weber, J. N., Griffiths, G. M., Tanaka, Y., Osame, M. & Bangham, C. R. M. 2003. Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. *Science*, **299,** 1713-1716.

Kaplan, J. E. & Khabbaz, R. F. 1993. The epidemiology of human T-cell leukemia virus types 1 and 2. *Rev. Med. Virol.*, **3,** 137-148.

Khabbaz, R. F., Darrow, W. W., Hartley, T. M., Witte, J., Cohen, J. B., French, J., Gill, P. S., Potterat, J., Sikes, R. K., Reich, R., Kaplan, J. E. & Lairmore, M. D. 1990. Seroprevalance and risk factors for HTLV-I/II infection among prostitutes in the United States. *J. Am. Med. Assoc.*, **263,** 60-64.

Lee, B. M., De Guzman, R. N., Turner, B. G., Tjandra, N. & Summers, M. F. 1998. Dynamic Behavior of the HIV-1 Nucleocapsid Protein. *J. M. B.*, **279,** 633-649.

Matsushita, K., Matsumoto, T., Ohtsubo, H., Fujiwara, H., Imamura, N., Hidaka, S., Kukita, T., Tei, C., Matsumoto, M. & Arima, N. 1999. Long term maintenence combinatorial chemotherapy with OPEC/MPEC (Vincristine or Methotrexate, Prednisolone, Etoposide and Cyclophosphamide) or with Daily Oral Etoposide and Prednisolone can improve survival and quality of life in adult T-cell leukemia/lymphoma. *Leuk. Lymphoma*, **36,** 67-75.

Murphy, R. A. & Kingsbury, D. W. 1990. Virus Taxonomy. In: *Fields Virology, 2nd Ed.* (Ed. by Knipe, B. N. F. a. D. M.), pp. 26-27. New York: Raven Press.

Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. & Gallo, R. C. 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A*, **77,** 7415-9.

Satcher, D. 1995. Emerging infections: Getting ahead of the curve. *Emerg. Infect. Dis.*, **1,** 1-6.

Stephen, A. G., Worthy, K. M., Towler, E., Mikovits, J. A., Sei, S., Roberts, P., Yang, Q., Akee, R. K., Klausmeyer, P., McCloud, T. G., Henderson, L., Rein, A., Covell, D. G., Currens, M., Shoemaker, R. H. & Fisher, R. J. 2002. Identification of HIV-1 Nucleocapsid Protein: Nucleic Acid Antagonists with Cellular Anti-HIV Activity. *Bioche. Biophys. Res. Commun.*, **296,** 1228-1237.

Williams, M. C., Gorelick, R. J. & Musier-Forsyth, K. 2002. Specific Zinc-finger Architecture Required for HIV-1 Nucleocapsid Protein's Nucleic Acid Chaperone Function. *P.N.A.S.*, **99,** 8614-8619.

Yoshida, M., Myoshi, I. & Hinuma, Y. 1982. Isolation and characterization of retrovirus from cell lines of human T-cell leukemia and its implication in the disease. *Proc. Natl. Acad. Sci. USA*, **79,** 2031.

Yoshida, M., Seiki, M., Yamaguchi, K. & Takatsuki, K. 1984. Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease. *Proc. Natl. Acad. Sci. USA*, **81,** 2534-2537.

Zucker-Franklin, D. & Pancake, B. B. 1998. HTLV-I tax among American blood donors. *Clin. Diagn. Lab. Immunol.*, **5,** 831-835.

CHAPTER 2

HTLV-I GENOMIC STRUCTURE

2.1 Viral Assembly

The transfer of an infectious virus requires the assembly of its structure. Proteins comprising the retrovirus are expressed from an RNA genome, via a DNA intermediate, as precursor polyproteins: Gag, Gag-Pro-Pol, and Env. These polyproteins are derived from the group-specific antigen (*gag*), the protease (*pro*), the polymerase (*pol*), and the envelope (*env*) genes on the RNA genome (Koralnik, 1996; Coffin et al., 1997). The expressed polyproteins are proteolytically cleaved to generate the internal and structural proteins. The spherical core proteins of oncoviruses or the conical core proteins of lentiviruses consist of the matrix (MA), capsid (CA), and nucleocapsid (NC), which are associated with the Gag precursor. The protease (PR) is associated with the Gag-Pro and Gag-Pro-Pol precursors and reverse transcriptase (RT) and integrase (IN) are associated only with the Gag-Pro-Pol precursor (Coffin et al., 1997). The envelope proteins of the surface (SU) and the transmembrane (TM) are associated with the Env. Figure 2.1 shows an illustration of the virion components of a retrovirus.

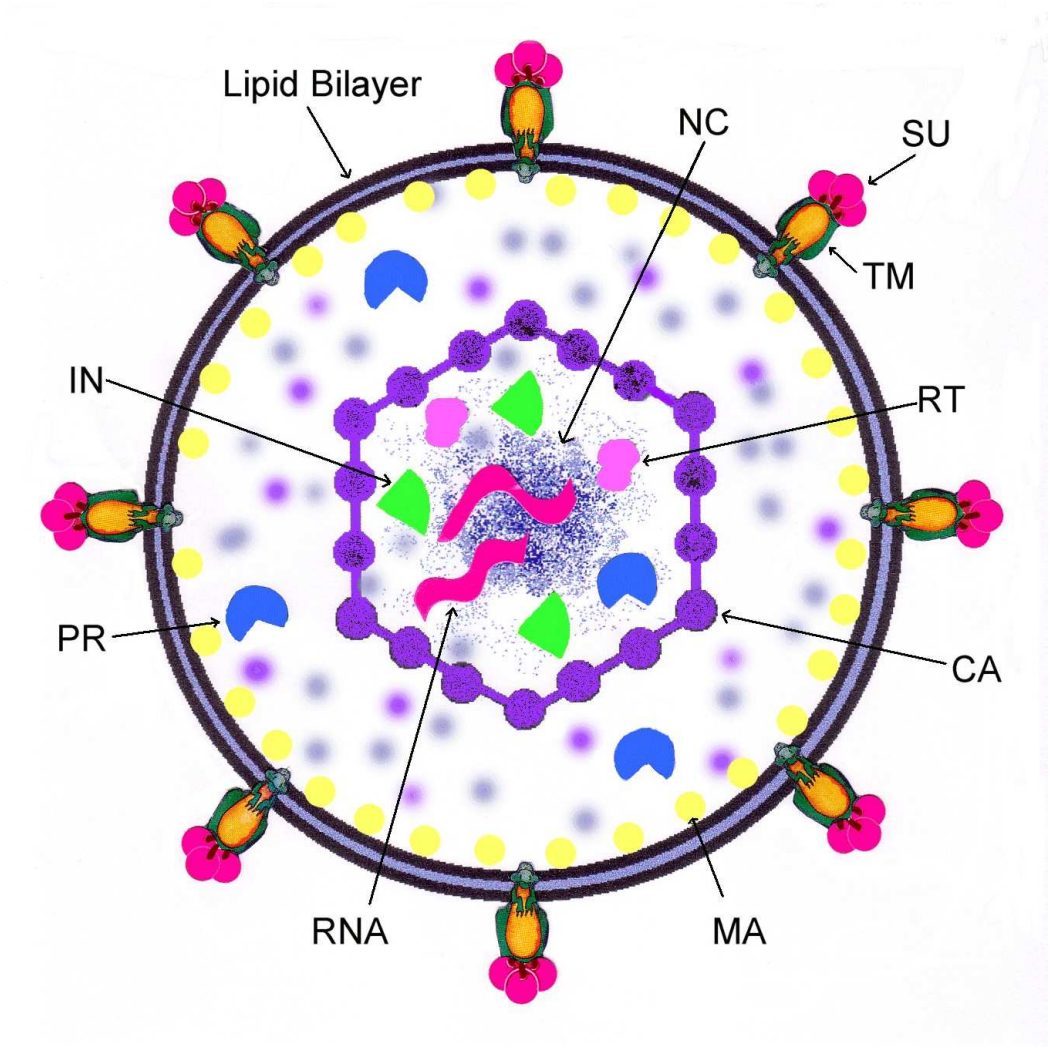


Figure 2.1 General retroviral protein components. Abstracted from Coffin *et al. Retroviruses*, 1997 (NC = nucleocapsid, Su = surface, TM = transmembrane, RT = reverse transcriptase, CA = capsid, MA = matrix, PR = protease, IN = integrase).

Figure 2.2 shows an illustration of typical retroviral assembly. Assembly begins with the viral transcripts used to synthesize the Gag and Gag-Pro-Pol polyproteins and ends with the immature budding virion followed by protease processing to form the mature virion (Coffin et al., 1997).

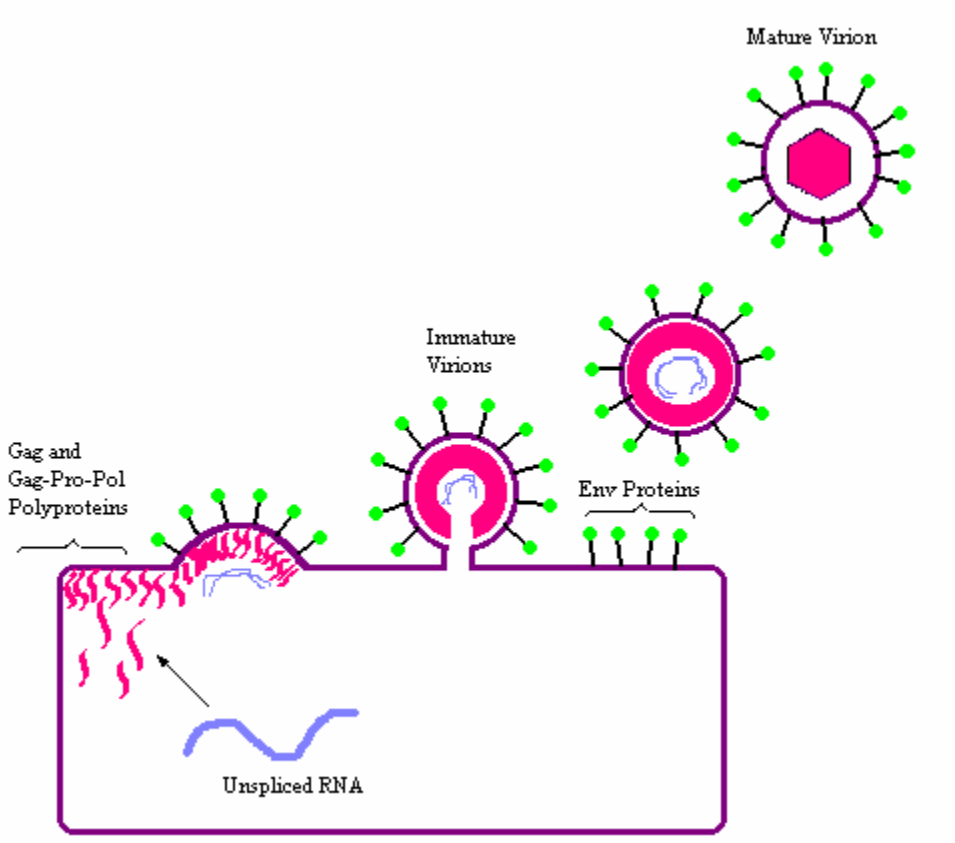


Figure 2.2 Typical retroviral assembly. Abstracted from Coffin *et al. Retroviruses*, 1997.

The budding process is facilitated by myristylation of the MA-end of the Gag, Gag-Pro, and Gag-Pro-Pol precursors. These precursors then migrate toward the host cell lipid bilayer. The unprocessed proteins are assembled into a spherical shell and begin the budding process (Craven & Parent, 1996).

2.2 HTLV-I Precursor Proteins

The proviral genome of HTLV-I is 9033 base pairs in length (Zhao et al., 1993) and has been isolated and cloned from the leukocytes of ATL patients. The subsequent characterization of the genome revealed that it is composed of *gag*, *pol*, *pro*, and *env*, genes as previously described (Kobayashi et al., 1991; Koralnik, 1996). Additionally, HTLV-I has two non-structural genes, *tax* and *rex* that code for regulatory proteins involved in synthesis and processing of the viral RNA (Murphy & Kingsbury, 1990). The genes are flanked by long terminal repeats (LTRs) which are divided into three sequences: U3, U5, and R (Figure 2.3). U3 and U5 relate to sequences at the 3' and 5' ends of the RNA genome and R relates to a sequence repeated at both 3' and 5' end (Coffin et al., 1997). The LTRs are formed in the 3' to 5' direction with U3-R-U5 during DNA synthesis. As illustrated in Figure 2.3, the proviral DNA is considerably longer than the original full-length viral RNA. The LTRs contain strong promoter and enhancer elements (Andersson, 2002; Coffin et al., 1997).

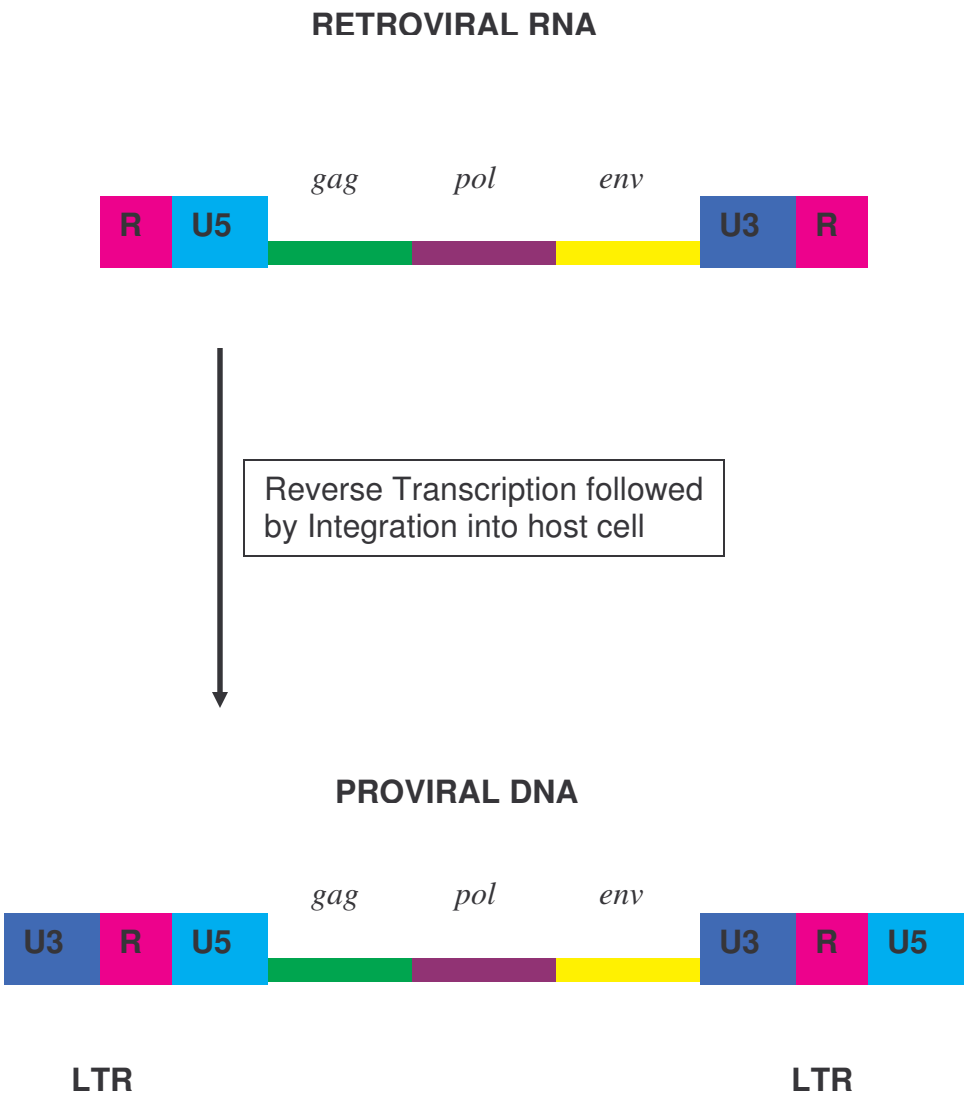


Figure 2.3 Retroviral gene ORFs with flanking LTRs.

The open reading frame (ORF) of the protease is located between the *gag* and *pol* genes. The *gag*, *pro*, *pol* organization of HTLV-I is different than that of HIV-1 (Figure 2.4). The ORF is shifted in the –1 direction of the *pro-pol* (*pro* is the 5' end of *pol*) from *gag* in HIV. In HTLV-I, the *gag*, *pro*, *pol* are separate genes encoded in different ORFs. In HTLV-I, the *pro* is frame shifted in the –1 direction from *gag* and *pol* is shifted in the –1 direction from *pro*. In HTLV-I, there are two ribosomal frame shifts for the *gag-pol* overlap and the *pro-pol* overlap that are necessary for synthesis of the *pol*-encoded HTLV-I replication enzymes (Ding et al., 1998). The 5' end of the *env* ORF overlaps with the 3' end of *pol* (Koralnik, 1996).

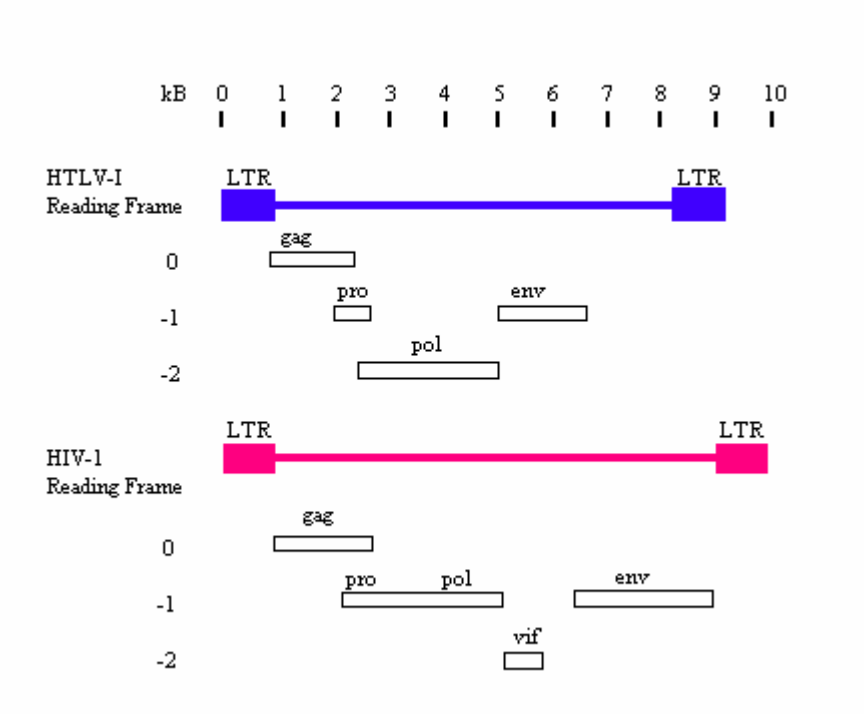


Figure 2.4 Schematic representation of the major genes for HTLV-I and HIV. The scale is in kilobases (kB) shown at the top of the diagram. The "Reading Frame" refers to the translation frame of the illustrated genes relative to *gag*, arbitrarily assigned as the 0 position. Note: Only major genes are shown. Adapted from Ding *et al.,*1998.

The occurrence of the two frame shifts is required for the replication of the virus (Nam et al., 1988). Translation of the genome yields four large fusion proteins: Gag, Gag-Pro, Gag-Pro-Pol, and Env (Nam et al., 1993; Carrington & Schulz, 1996; Koralnik, 1996). The protease is necessary for cleavage of these polyprecursor proteins into the mature products as shown in Figure 2.5.

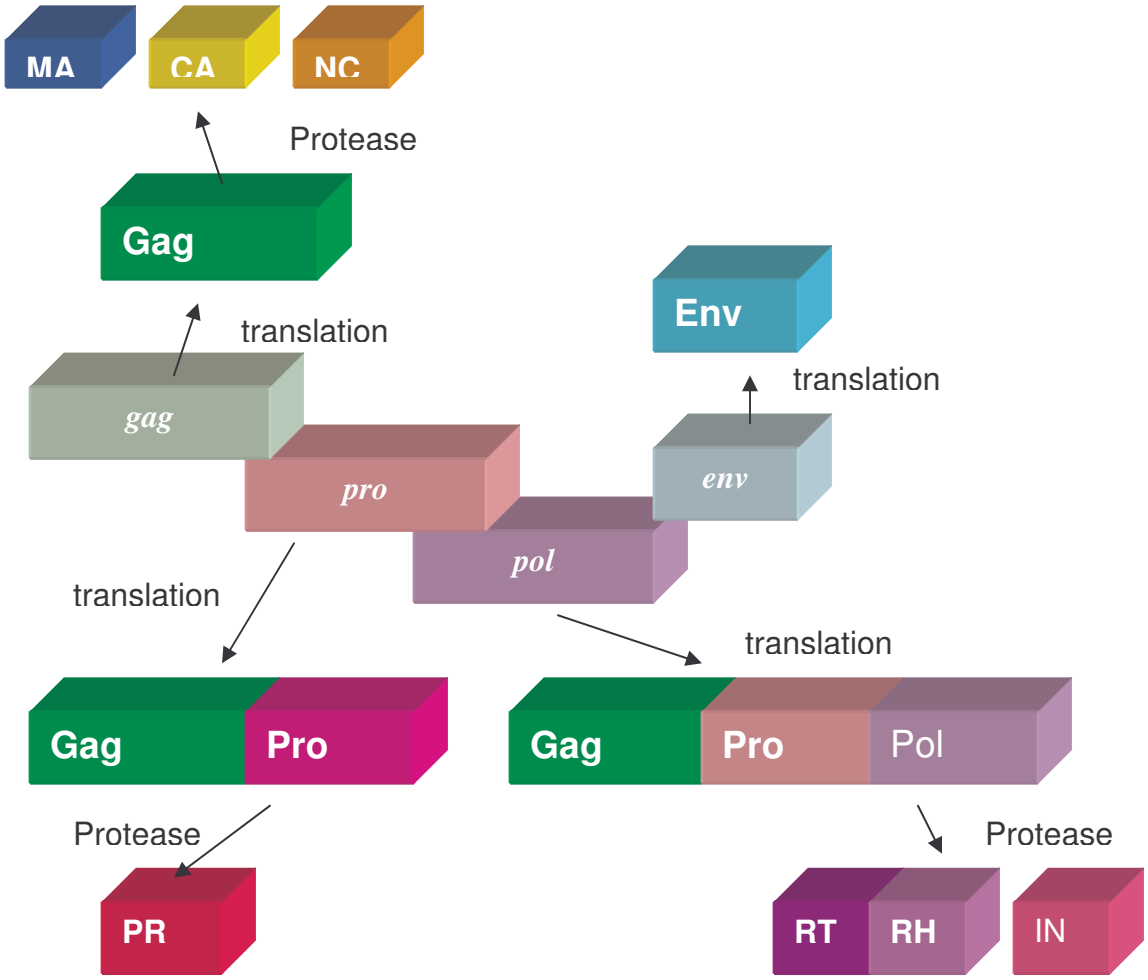


Figure 2.5 Products of genomic translation and processing.

Figure 2.6 shows the one-letter amino acid sequences of the Gag, Gag-Pro, and Gag-Pro-Pol fusion proteins prior to processing by HTLV-I protease.

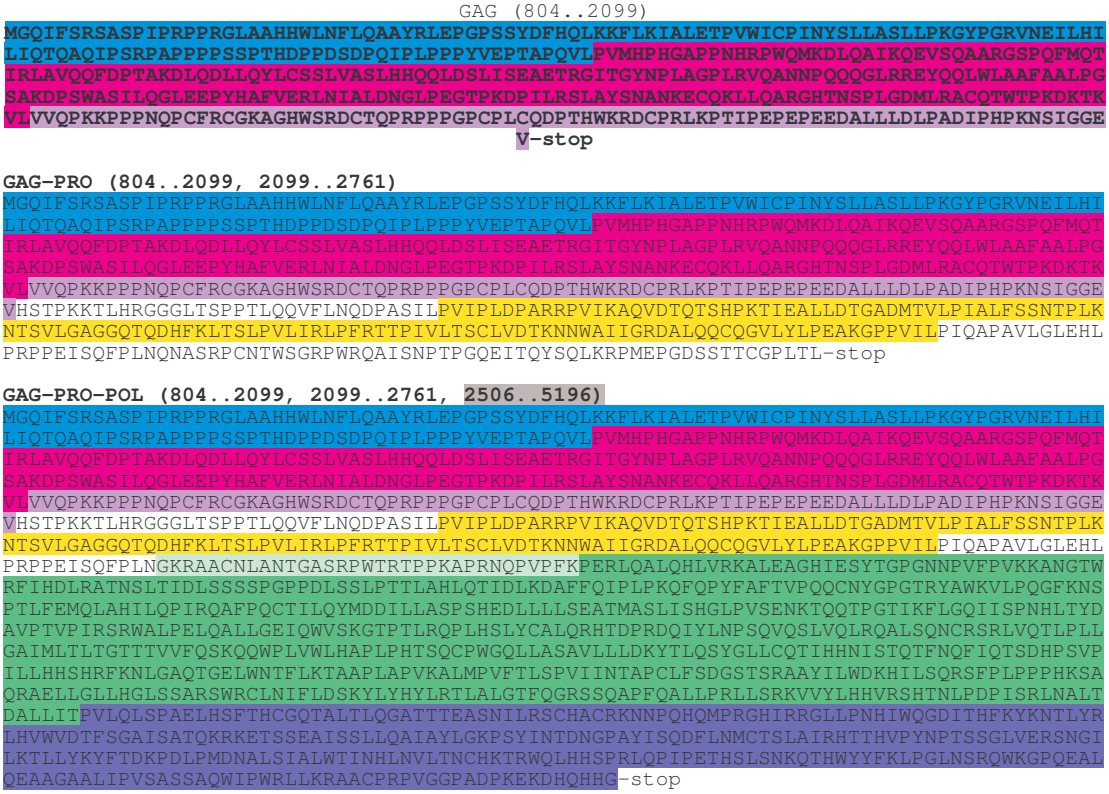


Figure 2.6 Gag, Gag-Pro, and Gag-Pro-Pol fusion proteins taken from HTLV-I K30 clone, GenBank Accession L03562. Colors indicate protein product sequences. Stops indicate where stop codons exist across the ORFs. Zhao *et al.* (1993) *Virology* **195**, 271-274.

Cleavage of the Gag protein by HTLV-I protease yields proteins: p19 MA (130 amino acids), p24 CA (214 amino acids), and p15 NC (85 amino acids) p19, p24, and p15 signify that these proteins are 19 kDa, 24 kDa, and 15 kDa, respectively (Koralnik,

1996). The protease is derived from Pro and proteolytic cleavage of the Pol results in the RT and IN production (Nam et al., 1993; Nam et al., 1988; Nam & Hatanaka, 1986). The Env precursor is glycosylated and then processed by exopeptidases to yield the 46 kDa SU and 21 kDa TM envelope glycoproteins (Coffin et al., 1997).

The exact mechanism for how the protease cleaves itself from the Gag-Pro-Pol and Gag-Pro has not been resolved. For HIV, it has been shown that the protease is active as a dimer (Coffin et al., 1997; Oroszlan & Luftig, 1990). The concentration of the Gag-Pol precursor protein may actually facilitate dimerization. Therefore, it is highly likely that initial proteolytic cleavage is carried out by the dimerization of two Gag-Pol precursor proteins within the immature protease (Coffin et al., 1997; Pettit et al., 2003). The development of protease inhibitors that disrupt this proteolytic processing is an attractive target and should be an effective strategy for rational drug design that will provide significant advancement in the treatment of HTLV-I related diseases.

REFERENCES

Andersson, A. 2002. Studies on human endogenous retroviruses (HERVs) with special focus on ERV3. In: *Comprehensive Summaries of Uppsala Dissertations from Faculty of Medicine 1165.* Uppsala, Sweden: Acta Unversitatis Upsaliensis.

Carrington, C. V. F. & Schulz, T. F. 1996. Virology of HTLV-I infection. In: *Human T-Cell Lymphotropic Virus Type I* (Ed. by Hafler, P. H. a. D. A.), pp. 111-139. Chichester, UK: Wiley.

Coffin, J. M., Hughes, S. H. & Varmus, H. E. 1997. *Retroviruses*. Plainview: Cold Spring Harbor Laboratory Press.

Craven, R. & Parent, L. 1996. The dynamic interactions of the Gag polyprotein. In: *Morphogenesis and Maturation of Retroviruses* (Ed. by Krausslich, H. G.), pp. 65-94.

Ding, Y. S., Owen, S. M., Lal, R. B. & Ikeda, R. A. 1998. Efficient expression and rapid purification of human T-cell leukemia virus type 1 protease. *J Virol*, **72,** 3383-6.

Kobayashi, M., Ohi, Y., Asano, T., Hayakawa, T., Kato, K., Kakinuma, A. & Hatanaka, M. 1991. Purification and characterization of human T-cell leukemia virus type I protease produced in Escherichia coli. *FEBS Lett*, **293,** 106-10.

Koralnik, I. J. 1996. Structure of HTLV-I. In: *Human T-cell Lymphotropic Virus Type I* (Ed. by Hollsberg, P. a. H., D. A.), pp. 65-78. Chichester, UK: Wiley.

Murphy, R. A. & Kingsbury, D. W. 1990. Virus Taxonomy. In: *Fields Virology, 2nd Ed.* (Ed. by Knipe, B. N. F. a. D. M.), pp. 26-27. New York: Raven Press.

Nam, S. H., Copeland, T. D., Hatanaka, M. & Oroszlan, S. 1993. Characterization of ribosomal frameshifting for expression of pol gene products of human T-cell leukemia virus type I. *J Virol*, **67,** 196-203.

Nam, S. H. & Hatanaka, M. 1986. Identification of a protease gene of human T-cell leukemia virus type I (HTLV-I) and its structural comparison. *Bioche. Biophys. Res. Commun.*, **139,** 129-135.

Nam, S. H., Kidokoro, M., Shida, H. & Hatanaka, M. 1988. Processing of Gag precursor polyprotein of hman T-cell leukemia virus type I by virus-encoded protease. *J. Virol.*, **62,** 3718-3728.

Oroszlan, S. & Luftig, R. B. 1990. Retroviral Proteinases. *Curr. Top. Microbiol. Immunol.*, **157,** 153-185.

Pettit, S. C., Gulnik, S., Everitt, L. & Kaplan, A. H. 2003. The Dimer Interfaces of Protease and Extra-Protease Domains Influence the Activation of Protease and the Specificity of GagPol Cleavage. *J. Virol.*, **77,** 366-374.

Zhao, T. M., Robinson, M. A., Sawasdikosol, S., Simpson, R. M. & Kindt, T. J. 1993. Variation in HTLV-I Sequences from Rabbit Cell Lines with Diverse *in vivo* Effects. *Virology*, **195,** 271-274.

CHAPTER 3

HTLV-I PROTEASE

3.1 Aspartic Acid Proteases

Retroviral proteases are aspartyl proteases. These proteases have two aspartic acid residues in the active site that participate in the hydrolysis of peptides (Figure 3.1). The catalytic mechanism follows general acid-base catalysis. It is widely accepted that the carboxyl group of one of the aspartic acid residues is protonated while the second one is deprotonated upon substrate binding. The general scheme of catalysis is shown in Figure 3.2. The first step of the mechanism involves two concerted proton transfers, which facilitate nucleophilic attack by a water molecule on the substrate carbonyl carbon to form a tetrahedral intermediate. In the third step, the aspartate residue is protonated by one of the hydroxyl group from the tetrahedral intermediate, and the amine to be cleaved accepts a proton from the other aspartic acid residue (Garrett & Grisham, 1995).

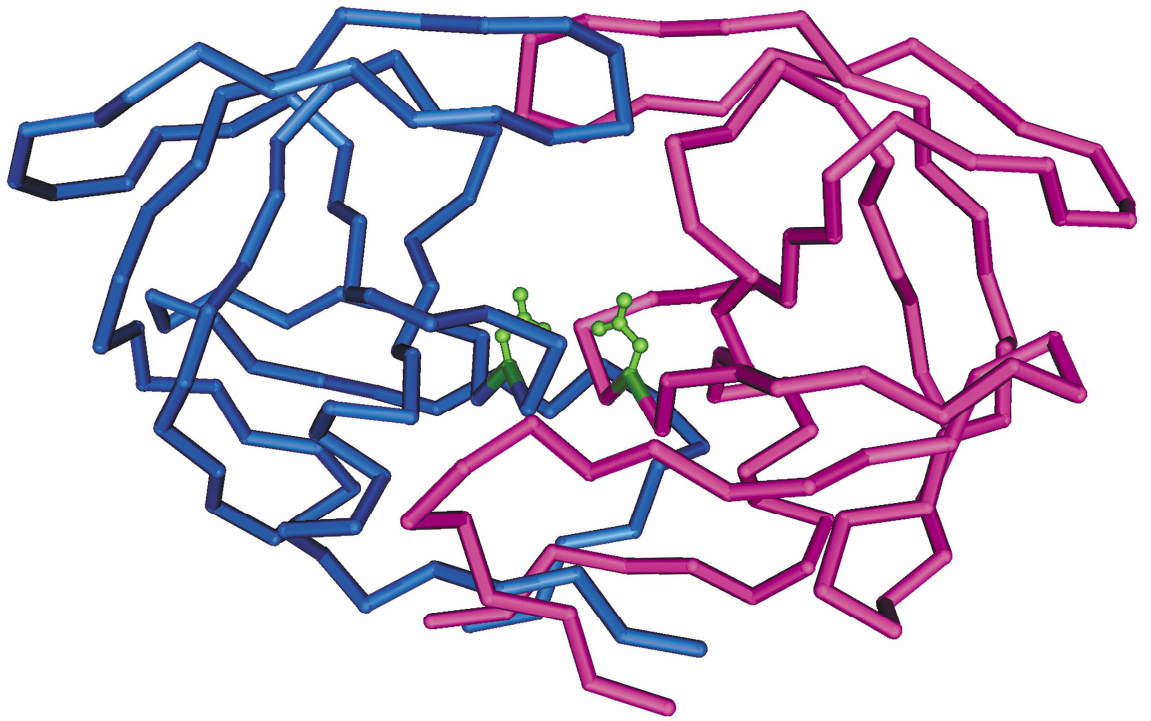


Figure 3.1 Structure of HIV-1 protease (PDB: 1A94). One monomer is colored in blue, the other in magenta. The catalytic aspartic acids are highlighted in green.

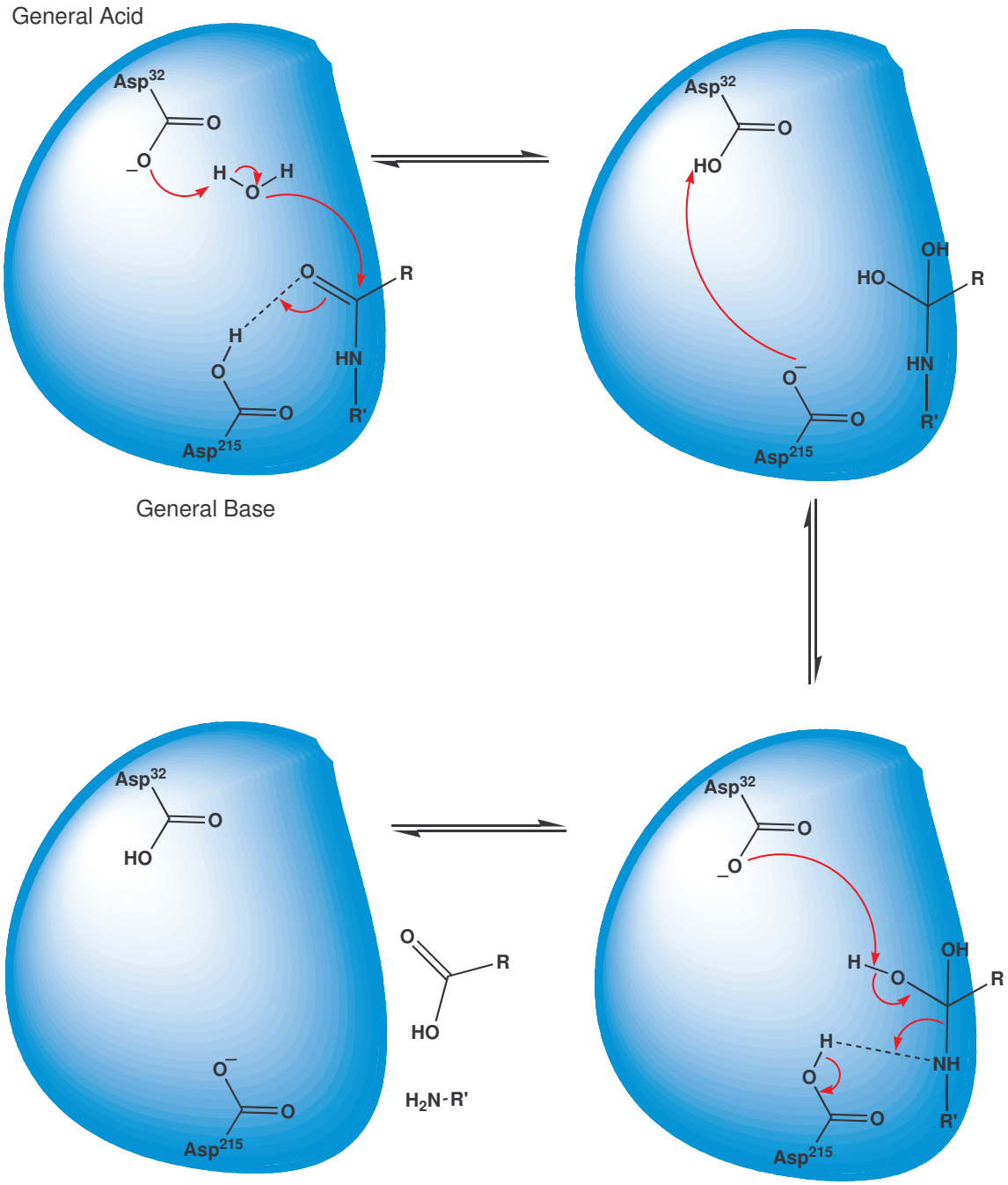


Figure 3.2 Mechanism of aspartic acid protease-catalyzed peptide cleavage.

3.2 HTLV-I Protease Homology

HTLV-I protease was isolated by Hatanaka in the late 1980's (Hatanaka & Nam, 1989) and by Daenke in the early 1990's (Daenke et al., 1994). HTLV-I protease is a homodimer aspartyl protease that has a conserved Asp-Thr-Gly catalytic site in the core active site region (Kobayashi et al., 1991). Several crystal structures have been determined for HIV-1, Rous Sarcoma Virus (RSV), and Feline Leukemia Virus (FIV) proteases (Gustchina et al., 1996; Louis et al., 1999). However, the structure of HTLV-I protease has not yet been determined. BLV protease has 32% identity with HTLV-I protease (Précigoux et al., 1993), making it the most homologous of all the retroviral proteases (Shuker et al., 2003). However, the three-dimensional structure of BLV protease has not been reported. Table 3.1 gives the homology percentages for known retroviral proteases.

Table 3.1 Homology of HTLV-I protease with other retroviral aspartic acid proteases. “Length” refers to the number of amino acids in the monomer.

	length	HTLV-I	HTLV-2	BLV	MuLV	RSV	HIV-1	HIV-2	SIV	FIV	EIAV
HTLV-I	125	-	50%	36%	21%	25%	28%	29%	32%	23%	26%
HTLV-2	125	50%	-	33%	2%	23%	26%	24%	26%	21%	25%
BLV	126	36%	33%	-	27%	23%	24%	25%	25%	21%	26%
MuLV	125	21%	25%	27%	-	21%	23%	24%	27%	25%	24%
RSV	124	25%	23%	23%	21%	-	30%	26%	25%	26%	26%
HIV-1	99	28%	26%	24%	23%	30%	-	48%	50%	25%	29%
HIV-2	99	29%	24%	25%	24%	26%	48%	-	86%	24%	33%
SIV	99	32%	26%	25%	27%	25%	50%	86%	-	22%	35%
FIV	113	23%	21%	21%	25%	26%	25%	24%	22%	-	30%
EIAV	104	26%	25%	26%	24%	26%	29%	33%	35%	30%	-

Figure 3.3 shows a sequence alignment with several aspartyl proteases: HTLV-I, BLV, RSV, HIV-1, HIV-2, and FIV. It is worth noting that the HTLV proteases have a long hydrophobic C-terminal “tail” that is only present in leukemia viral proteases (Herger, 2004; Shuker et al., 2003). Each chain of HTLV-I protease has 125 amino acid residues (115 without the “tail”), BLV has 126 amino acids (116 without the “tail”), RSV has 124 amino acids, and HIV has 99 amino acids (Shuker et al., 2003).

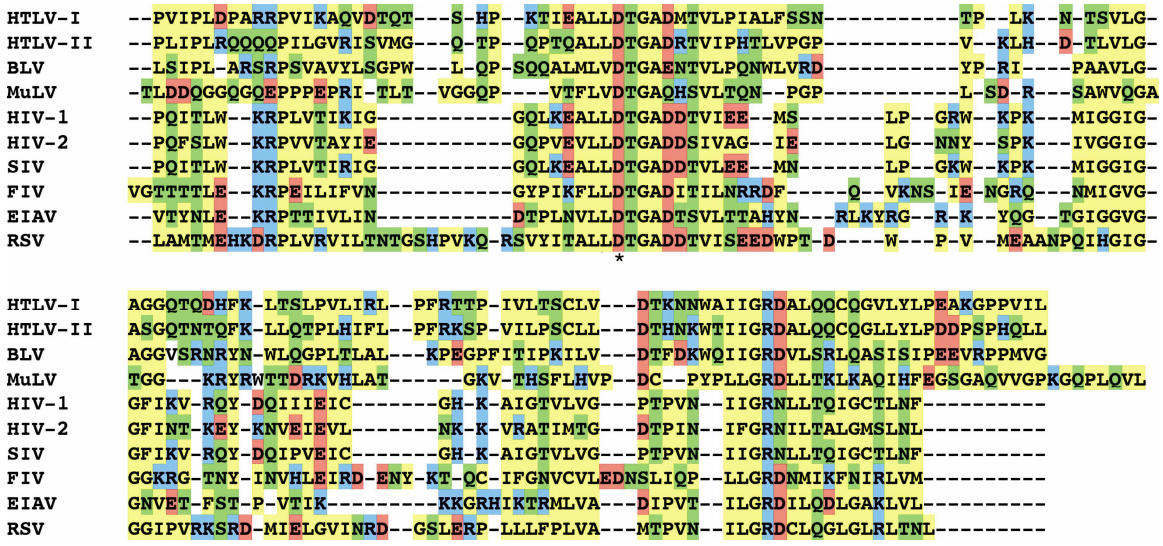


Figure 3.3 Homology sequence alignment of several aspartyl proteases.

Figure 3.4 shows a comparison of the backbones of the known HIV-1, HIV-2, and RSV protease structures superimposed. As can be seen in the Illustration, the backbone folds of the proteases are highly conserved.

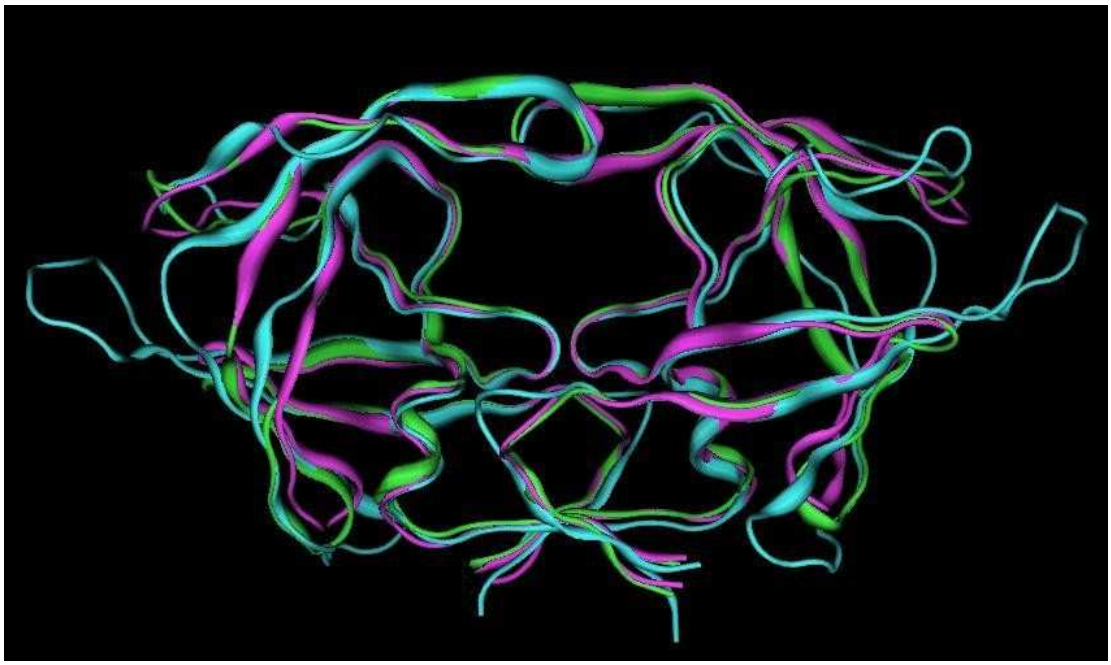


Figure 3.4 Superimposed backbone structures of HIV-1 (magenta), HIV-2 (green), and RSV (cyan) aspartyl proteases.

For proteins with no NMR or X-ray crystal structure, molecular modeling can be a very useful tool for understanding the three-dimensional structure. A good theoretical model should be determined by using a template with high resolution, minimal disorder, and a highly complete data collection set. In finding a suitable template structure for modeling HTLV-I protease, it is necessary to find a sequence with as much homology as possible. HIV-1 protease has less than 30% homology. In addition, the amino acid sequence differences between HIV-1 and HTLV-I proteases have been shown to produce differences in the activities of the enzymes; there are different specificities exhibited for protease inhibitors (Daenke et al., 1994; Hruskova-Heidingsfeldova et al., 1997; Kobayashi et al., 1991).

Crystal structures available as templates for HTLV-I protease are limited to RSV and HIV-1 protease structures deposited in the Protein Data Bank, which have low levels of homology. RSV protease bears a 21% sequence homology with HTLV-I protease, while HIV-1 protease is 28% homologous. However, RSV protease does have a region in the upper flap (RSV sequence: Gly-Ser-His-Pro-Val-Lys-Gln-Arg-Ser-Val) that is similar in number of residues to the corresponding HTLV-I protease region (HTLV-I sequence: Thr-Ser-His-Pro-Lys). These sequences are underlined with asterisks in Figure 3.5. This additional sequence of amino acids is absent in HIV-1 protease. Figure 3.5 also shows the extended outer loop residues (underlined with circles) of HTLV-I and RSV compared to the missing residues in HIV. Amino acides in color are the conserved homologous amino acids of RSV and HIV proteases compared with HTLV-I.

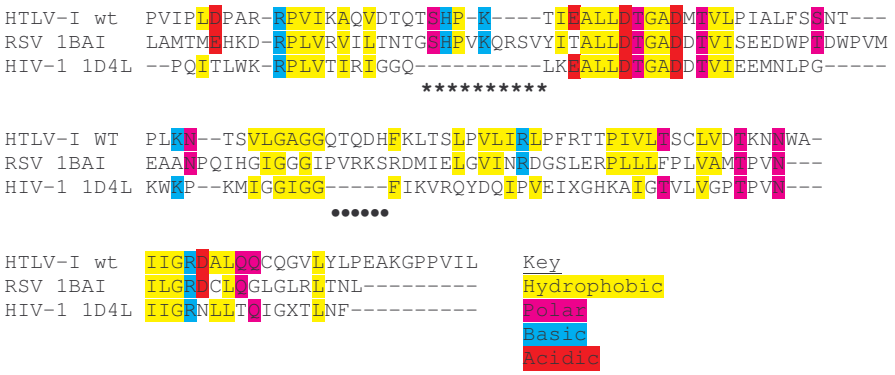


Figure 3.5 Sequence alignment of HTLV-I, RSV, and HIV proteases. Asterisks indicate the upper flap region and circles indicate the outer loop region.

In addition, it can be seen in Figure 3.5 that part of the upper flap region (RSV sequence: Pro-Val-Arg-Lys-Ser and HTLV-I sequence: Gln-Thr-Gln-Asp-His) is absent in HIV protease. Figure 3.6 shows a secondary structure map of RSV protease, illustrating

where the β-sheets and α-helices are located relative to the residue positions. This cartoon demonstrates typical retroviral protease secondary structure.

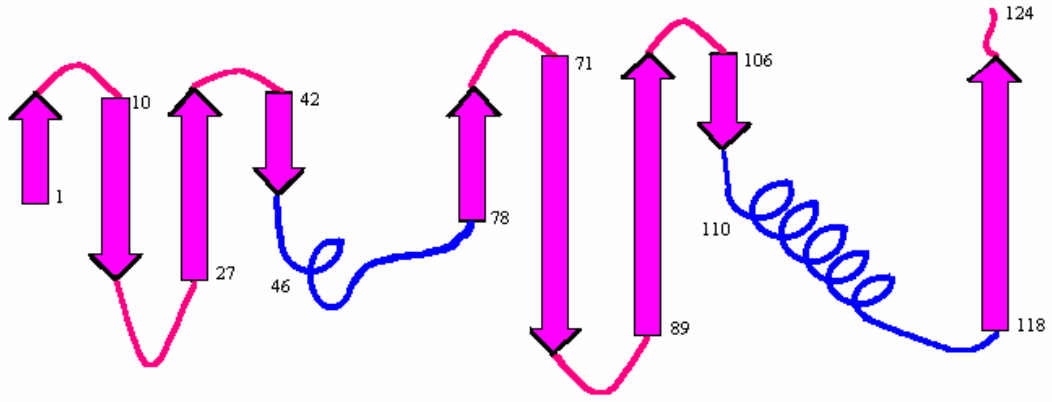


Figure 3.6 Schematic of secondary structural elements in retroviral proteases. Magenta arrows indicate β-sheets, blue coils are α-helices, and pink are random coils. Abstracted from Jaskólski *et al*. *Biochemistry* 1990, 29, 5893.

3.3 Theoretical Model of HTLV-I Protease

A theoretical model of HTLV-I protease (PDB 1O0J) (Herger, 2004) (Dennison et al., 2003) was generated using the program Modeller (Fiser et al., 2000; Marti-Renom et al., 2000; Sali & Blundell, 1993). This structure was modeled with RSV protease (PDB 1BAI) (Figure 3.7) serving as a structural template (Tözsér et al., 2000). This template

was chosen based on the quality of the published structure, as well as homology. For the crystal structure, RSV protease was crystallized with an HIV-1 protease peptidomimetic CA/p2 analog (Arg-Val-Leu-Nle-Phe-Glu-Ala-NH_2, Nle = norleucine, a reduced peptide.)

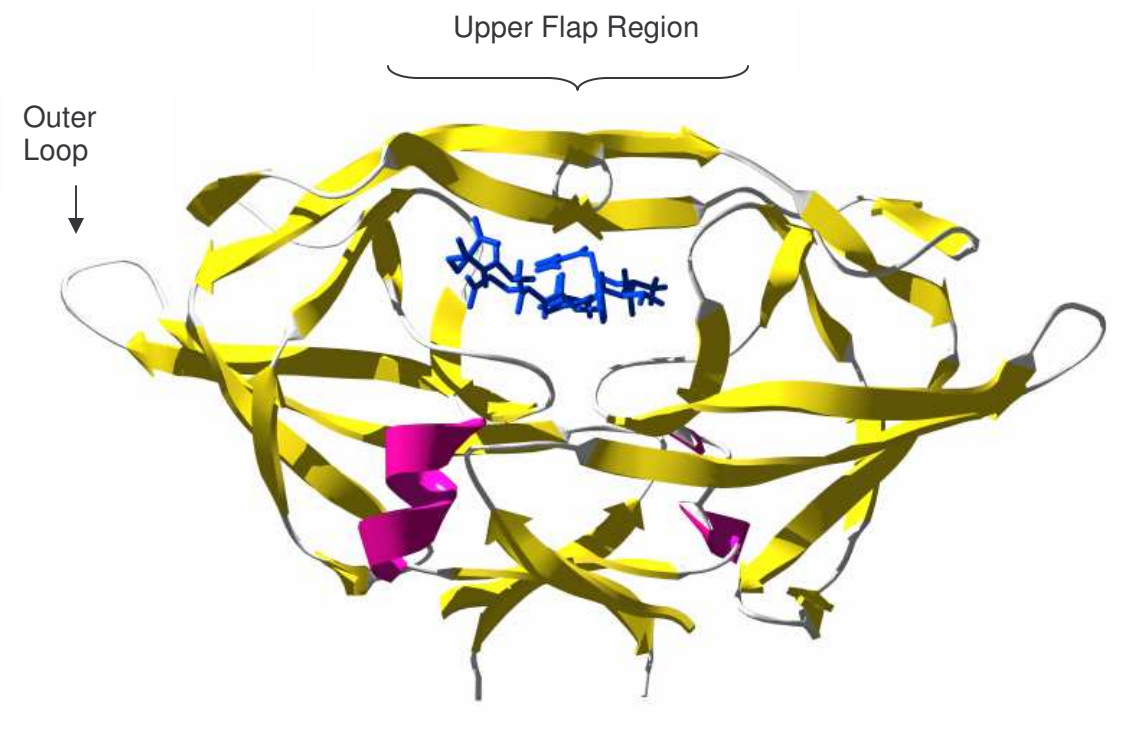


Figure 3.7 RSV protease (PDB 1BAI) with ligand (RVL-Nle-FEA-NH_2 shown as blue sticks) (Wu *et al*. 1998).

The RSV protease structure used as the template was refined at 2.4 Å. This crystal had 9 mutations (S38T, I42D, I44V, M73V, A100L, V104T, R105P, G106V, and S107N) to mimic HIV-1 protease. The data set was 76.8% complete. The final structure

had one dimer and one inhibitor in the asymmetric unit (Mahalingam et al., 1999; Weber et al., 1997; Wu et al., 1998).

Docking of the CA/NC (TKVLVVQP) HTLV-I substrate was accomplished by superimposing the MODELLER structure prepared by Bryan Herger over the RSV/RSV ligand complex using Molecular Operating Environment (MOE) (Chemical Computing Group). The RSV substrate was then mutated into the CA/NC HTLV-I substrate and subsequently minimized. This theoretical model (PDB 1O0J), based on our previous alignment, is shown in Figure 3.8 (Herger et al., 2004) and the superimposed theoretical model with HIV and RSV is shown in Figure 3.9. It is important to note that HTLV-I protease has an additional ten C-terminal residues that are not present in RSV protease and are therefore not included in the published PDB 1O0J structure.

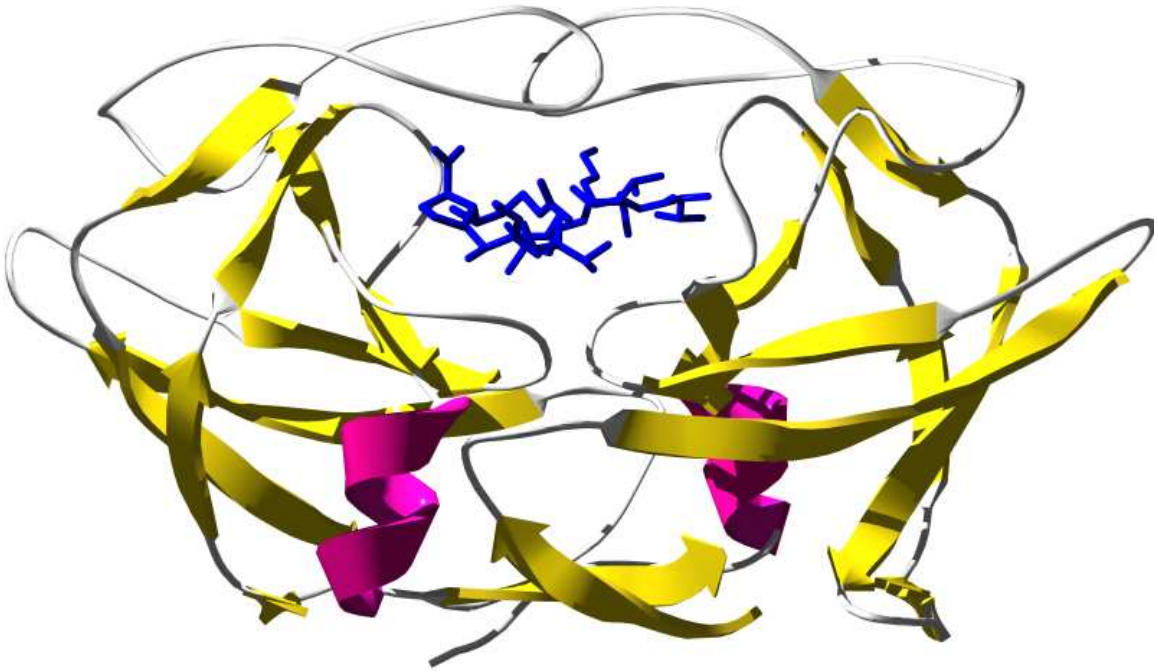


Figure 3.8 Theoretical model of HTLV-I protease (PDB 1O0J) with CA/NC substrate (Dennison, Herger, Shuker, 2003).

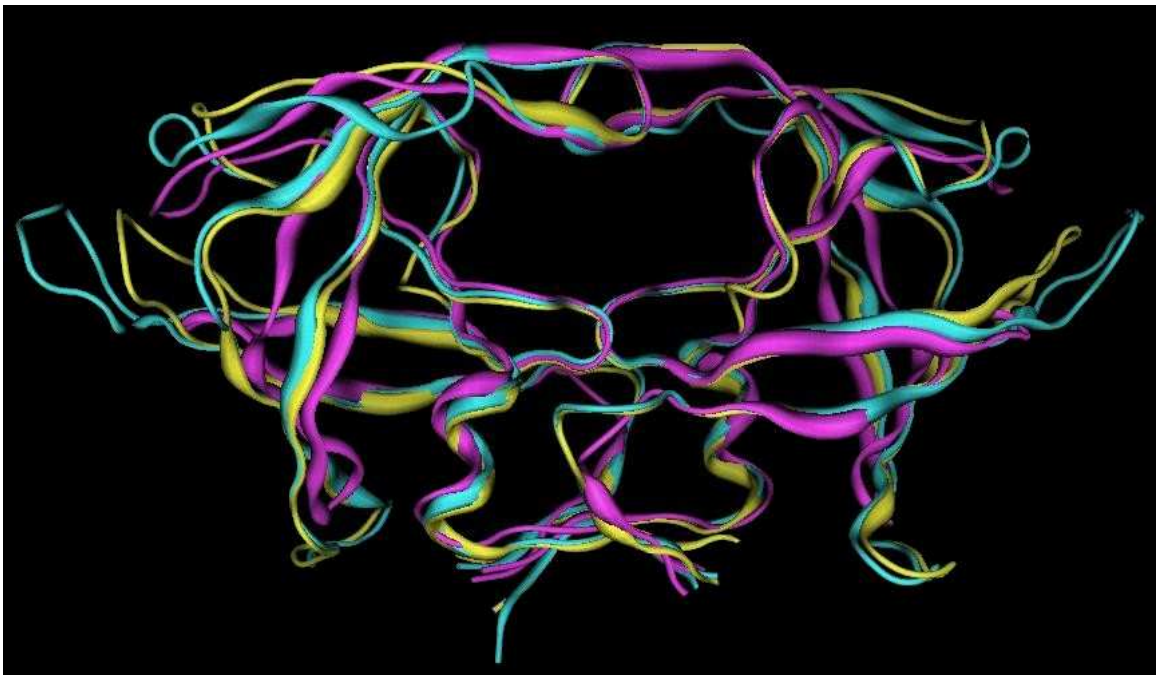


Figure 3.9 Superimposed backbone structures of theoretical HTLV-I protease model (gold), HIV-1 (magenta), and RSV (cyan) aspartyl proteases.

3.4 HTLV-I Protease Substrates

HTLV-I protease cleaves long chains of amino acids (the polyfusion proteins or precursor proteins) encoded by the *gag, gag-pro*, and *gag-pro-pol,* genes. There are several specific cleavage points along these chains of amino acids (called the scissile bonds) that bind tightly to HTLV-I protease as natural substrates. Figure 2.6 shows an illustration of the transcribed Gag, Gag-Pro, and Gag-Pro-Pol fusion proteins prior to proteolysis by HTVL-I protease. These substrates are then cleaved into smaller chains

of amino acids that form the MA, CA, NC, RT-RH, IN, and PR proteins (Coffin et al., 1997). Some of these specific cleavage site substrates are shown in Table 3.2.

There are eight binding pockets in the active site of retroviral proteases. From scissile site moving toward the right, the pocket is numbered S1', S2', S3', S4'…and the substrate is numbered P1', P2', P3', P4'…From the scissile site moving toward the left, the pocket is numbered S1, S2, S3, S4…and the substrate is numbered P1, P2, P3, P4…as shown in Figure 3.10.

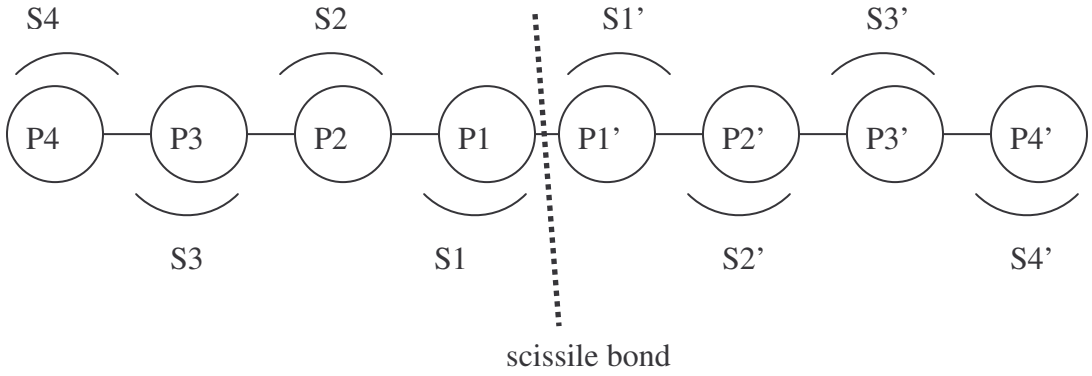


Figure 3.10 Schematic diagram of substrate positions and binding pockets. Circles represent substrate amino acids, and arcs represent sites on the protease. The scissile bond is the point of cleavage.

Table 3.2 Retroviral protease cleavage sites. The scissile bond is indicated by the slash.

Cleavage Junction	HTLV-I	BLV	MuLV	RSV	HIV-1	FIV	EIAV
MA/CA	PQVL/PVMH	PAIL/PIIS	--	VVAM/PVVI	SQNY/PIVQ	PQAY/PIQT	SEEY/PIMI
MA/pX	--	--	SSLY/PALT	--	--	--	--
X/CA	--	--	SQAF/PLRA	--	--	--	--
CA/NC	TKVL/VVQP	PAIL/VHTP	SKLL/ATVV	AAAM/SSIA	--	KMQL/LAEL	--
CA/pX	--	--	--	--	ARVL/AEAM	--	KMML/LAKA
pX/NC	--	--	--	--	ATIM/MQRG	--	AKAL/QTGL
NC/pX	--	--	--	--	RQAN/FLGK	--	--
Gag/PR	ASIL/PVIP	LECL/LSIP	TSLL/TLDD	PAVS/LAMT	SFNF/PQIT	GFVN/YNKV	QFVG/VTYN
PR/Pol	PVIL/PIQA	PMVG/VLDA	LQVL/TLNL	LTNL/IGRA	TLNF/PISP	RLVM/AIQS	KLVL/AQLS
TF1/RT	PAVL/GLEH	--	--	--	--	--	--
RT-RH/IN	VLQL/SPAD	unk	unk	FQAY/PLRE	RKIL/FLDG	CQTM/MIIE	TGVF/WVEN

In HTLV-I protease, the S4 and S4' subsites are shallow, surface exposed pockets. In the natural substrates, the P4 and P4' positions are typically hydrophobic residues, with proline being the most prevalent. A variety of substrates are tolerated, which includes polar residues such as serine and threonine and charged residues such as aspartic acid and histidine.

Unlike the S4 and S4' subsites, the S3 and S3' pockets are large and can accommodate a wide variety of side chains. The natural substrate of CA/NC has a lysine residue in the P3 position. However, a hydrophobic residue such as phenylalanine, leucine, or valine in this position is catalyzed with nearly the same efficiency. The S2 and S2' subsites are also large and are particularly inhabited by the β-branched residues of valine and isoleucine. The S1 subsite in HTLV-I protease is occupied by a P1 leucine. In this site, a P1 β-branched amino acid is not tolerated. The S1' typically is occupied by a P1' proline. Therefore, the most common scissile site (P1/P1') is Leu/Pro. Table 3.3 shows the natural substrate cleavage sites for HTLV-I protease.

Table 3.3 HTLV-I Protease Cleavage of Natural Substrates.

JUNCTION	CLEAVAGE SITE
MA/CA	APQVL/PVMHP
CA/NC	KTKVL/VVQPK
TF1/PR	PASIL/PVIPL
PR/TFP	PPVIL/PIQAP
TFP/RT	APAVL/GLEHL
RT-RH/IN	PVLQL/SPADL

As can be seen in Table 3.3, all the P1 residues are leucine. The P1' residues are mostly proline with the exception of a P1' valine in the CA/NC substrate, a recently reported P1' glycine in the TFP/RT substrate (Heidecker et al., 2002), and an unusual P1' serine (Mariani & Shuker, 2003; Shuker et al., 2003).

REFERENCES

Coffin, J. M., Hughes, S. H. & Varmus, H. E. 1997. *Retroviruses*. Plainview: Cold Spring Harbor Laboratory Press.

Daenke, S., Schramm, H. J. & Bangham, C. R. 1994. Analysis of substrate cleavage by recombinant protease of human T cell leukaemia virus type 1 reveals preferences and specificity of binding. *J Gen Virol*, **75,** 2233-9.

Dennison, K. J., Herger, B. E. & Shuker, S. B. 2003. Protein Data Bank ID 1O0J.

Fiser, A., Do, R. K. & Sali, A. 2000. Modeling of loops in protein structures. *Protein Sci*, **9,** 1753-73.

Garrett, R. H. & Grisham, C. M. 1995. *Biochemistry*. Orlando: Saunders College Publishing.

Gustchina, A., Kervinen, J., Powell, D. J., Zdanov, A., Kay, J. & Wlodawer, A. 1996. Structure of equine infectious anemia virus proteinase complexed with an inhibitor. *Protein Sci*, **5,** 1453-65.

Hatanaka, M. & Nam, S. H. 1989. Identification of HTLV-I gag protease and its sequential processing of the gag gene product. *J. Cell. Biochem.*, **40,** 15-30.

Heidecker, G., Hill, S., Lloyd, P. A. & Derse, D. 2002. A novel protease processing site in the transframe protein of human t-cell leukemia virus type 1 $PR76^{gag\text{-}pro}$ defines the N terminus of RT. *J. Virol.*, **76,** 13101-13105.

Herger, B. E., Mariani, V. M., Dennison, K. J., & Shuker, S. B. 2004. The Ten C-Terminal Residues of HTLV-I Protease Are Not Necessary for Enzymatic Activity. *Biochemical and Biophysical Research Communications*, **In press**.

Hruskova-Heidingsfeldova, O., Blaha, I., Urban, J., Strop, P. & Pichova, I. 1997. Substrates and inhibitors of human T-cell leukemia virus type 1 (HTLV- 1) proteinase. *Leukemia*, **11 Suppl 3,** 45-6.

Kobayashi, M., Ohi, Y., Asano, T., Hayakawa, T., Kato, K., Kakinuma, A. & Hatanaka, M. 1991. Purification and characterization of human T-cell leukemia virus type I protease produced in Escherichia coli. *FEBS Lett*, **293,** 106-10.

Louis, J. M., Oroszlan, S. & Tözsér, J. 1999. Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type I proteinase. *J. Biol. Chem.*, **274,** 6660-6666.

Mahalingam, B., Louis, J. M., Reed, C. C., Adomat, J. M., Krouse, J., Wang, Y. F., Harrison, R. W. & Weber, I. T. 1999. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur J Biochem*, **263,** 238-45.

Mariani, V. M. & Shuker, S. B. 2003. Identification of the RT-RH/IN cleavage site of HTLV-I. *Bioche. Biophys. Res. Commun.*, **300,** 268-270.

Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, **29,** 291-325.

Précigoux, G., Geoffre, S., Leonard, R., Llido, S., Dautant, A., Destaintot, B. I., Picard, P., Menard, A., Guillemain, B. & Hospital, M. 1993. Modelling, synthesis and biological activity of a BLV proteinase, made of (only) 116 amino-acids. *FEBS Lett.*, **326,** 237-240.

Sali, A. & Blundell, T. L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, **234,** 779-815.

Shuker, S. B., Mariani, V. M., Herger, B. E. & Dennison, K. J. 2003. Understanding HTLV-I Protease. *Chem. Biol.*, **5,** 373-380.

Tözsér, J., Zahuczky, G., Bagossi, P., Louis, J. M., Copeland, T. D., Oroszlan, S., Harrison, R. W. & Weber, I. T. 2000. Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. *Eur. J. Biochem.*, **267,** 6287-6295.

Weber, I. T., Wu, J., Adomat, J., Harrison, R. W., Kimmel, A. R., Wondrak, E. M. & Louis, J. M. 1997. Crystallographic analysis of human immunodeficiency virus 1 protease with an analog of the conserved CA-p2 substrate -- interactions with frequently occurring glutamic acid residue at P2' position of substrates. *Eur J Biochem*, **249,** 523-30.

Wu, J., Adomat, J. M., Ridky, T. W., Louis, J. M., Leis, J., Harrison, R. W. & Weber, I. T. 1998. Structural basis for specificity of retroviral proteases. *Biochemistry*, **37,** 4518-26.

CHAPTER 4

HTVL-I PROTEASE C-TERMINAL CLEAVAGE

4.1 HTLV-I C-Terminal Cleavage Junction

As can be seen in the sequence alignment in Figure 4.1, HTLV-I protease has an extended C-terminal "tail" that is not common to other retroviral proteases with the exception of other leukemia viral proteases (Shuker et al., 2003).

```
HTLV-I   ---IGRDALQQCQGVLYLPEAKGPPVIL
HTLV-II  ---IGRDALQQCQGLLYLPDDPSPHQLL
BLV      ---IGRDVLSRLQASISIPEEVRPPMVG
MuLV     L--LGRDLLTKLKAQIHFEGSGAQVVGPKGQPLQVL
RSV      ---LGRDCLQGLGLRLTNL---------
HIV-1    ---IGRNLLTQIGCTLNF----------
HIV-2    ---FGRNILTALGMSLNL----------
SIV      ---FGRNLLTALGMSLNL----------
```

Figure 4.1 Sequence alignment of C-terminal regions of several retroviral proteases. BLV, bovine leukemia virus; MuLV, murine leukemia virus; RSV, Rous sarcoma virus; HIV, human immunodeficiency virus 1 and 2; SIV, simian immunodeficiency virus. Potential cleavage sites in HTLV proteases highlighted in yellow.

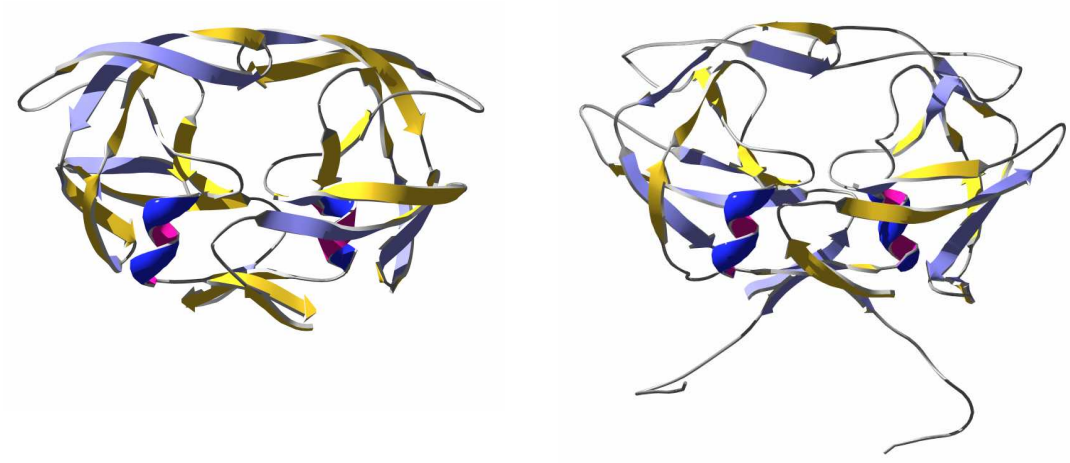
Within this tail domain, there is a potential cleavage site between residues Leu^{115}-Pro^{116} at the N-terminus of the last 10 amino acids (Figure 4.1), with a corresponding cleavage site in HTLV-II.

Several natural amino acid sequences recognized by HTLV-I protease have been reported (Kobayashi et al., 1991; Heidecker et al., 2002; Mariani & Shuker, 2003; Shuker et al., 2003); these are shown again in Table 4.1. All of these sequences contain leucine in the P1 position, with proline as the most common residue in the P1' position.

Table 4.1 HTLV-I Protease Cleavage of its Natural HTLV-I Substrates (see also Figure 2.6).

JUNCTION	CLEAVAGE SITE
MA/CA	APQVL/PVMHP
CA/NC	KTKVL/VVQPK
TF1/PR	PASIL/PVIPL
PR/TFP	PPVIL/PIQAP
TFP/RT	APAVL/GLEHL
RT-RH/IN	PVLQL/SPADL

As there is no tail in the RSV protein on which the theoretical structure is based, it may be expected that the modeling results would produce an open, random coil structure for the tail as is shown in Figure 4.2(b). For comparison, HIV-1, which like RSV protease also has no tail, is shown in Figure 4.2(a). Whether or not this tail is important in the function of the protease is still uncertain.



(a) (b)

Figure 4.2 (a) Crystal structure of HIV protease (PDB 1D4L); (b) theoretical model of HTLV-I protease (PDB 1O0J).

A mutant of bovine leukemia virus (BLV) protease that does not contain its ten C-terminal residues was prepared by Précigoux and co-workers, and it correctly processes a peptide corresponding to the BLV MA/CA cleavage junction (Précigoux et al., 1993). On the other hand, it has been reported that the last ten C-terminal residues of HTLV-I protease are required for activity (Hayakawa et al., 1992). However, this was based on detection of unprocessed Gag during expression of deletion mutants, which were not purified or tested further (Herger et al., 2004) Victoria Mariani and Bryan Herger engineered and expressed a construct of HTLV-I protease (L40I) to prevent autoproteolysis based on work by Louis and co-workers and a construct (L40I Δ10) without the last ten C-terminal residues. In agreement with Précigoux's results for BLV, it was demonstrated that the truncated HTLV-I protease cleaves the HTLV-I MA/CA junction with similar catalytic efficiency to the native protease (Table 4.2).

Table 4.2 Catalytic Activity of HTLV-I Protease Constructs.

HTLV-I Protease Construct	k_{cat}/K_M ($M^{-1}s^{-1}$)
L40I	590 ± 50
L40I Δ10	680 ± 50

Since enzymatic activity appears to be unaffected by the "C-tail" region, investigation as to why this C-terminal tail is not cleaved when there is a natural Leu-Pro scissile site could be significant in understanding HTLV-I protease enzymology. Residues 114-117 are YLPE, sequentially. The common native pattern spanning the cleavage junction, P2 through P2', is to have a P2 valine or isoleucine; a P1 leucine; a P1' proline, valine, serine, or glycine; and a P2' valine or isoleucine. It is conceivable that point mutations led to the modification of one or both residues, resulting in a shift of the site further downstream in the TFP fragment. We hypothesized that the site was inhibited by either the steric effects of the P2 tyrosine or by the electronic effects of the P2' glutamic acid.

4.2 C-Terminal Tail Cleavage Junctions

Steric hindrance plays a significant role in HTLV-I protease's affinity for substrate, as demonstrated by the work of Louis and co-workers to analyze a second LP site (Leu40-Pro41) in HTLV-I protease. They report that mutating Leu^{40} to isoleucine eliminates autoproteolysis (Louis et al., 1999). Isoleucine differs from leucine only in the placement of the branch on the side chain. Further, the list of known substrates shows a clear preference for a small, aliphatic residue in the P2 position. The Leu^{115}-Pro^{116} junction, shown in Figure 4.1, is a putative cleavage site; however, as previously noted, this particular site is not cleaved during proteolytic processing of the protease.

In order to determine whether the C-terminal region is inhibited by steric interference of the P2 Tyr^{114} or the electronic effects of the charged P2' Glu^{117}, we incubated the wild-type HTLV-I protease with synthetic peptides bearing mutations in the P2, P2', and both substrate positions. The peptides VLVLPEAKGP, VLYLPEAKGP, and VLVLPVAK correspond to Y114V, E117V, and a double mutant with both Y114V and E117V, respectively. We also incubated a synthetic wild-type substrate (VLYLPEAKGP) with the protease. Cleavage products were analyzed by LC-MS. As expected, the wild-type substrate was not cleaved by the protease (data not shown). The LC-MS data for the VLVYLPEAKGP (Y114V) peptide (Figure 4.3) reveal only the parent ions at 1022.7 Da (1022.2 predicted) indicating that the Y114V mutant is not cleaved and Tyr^{114} does not play a role in preventing the C-terminal cleavage.

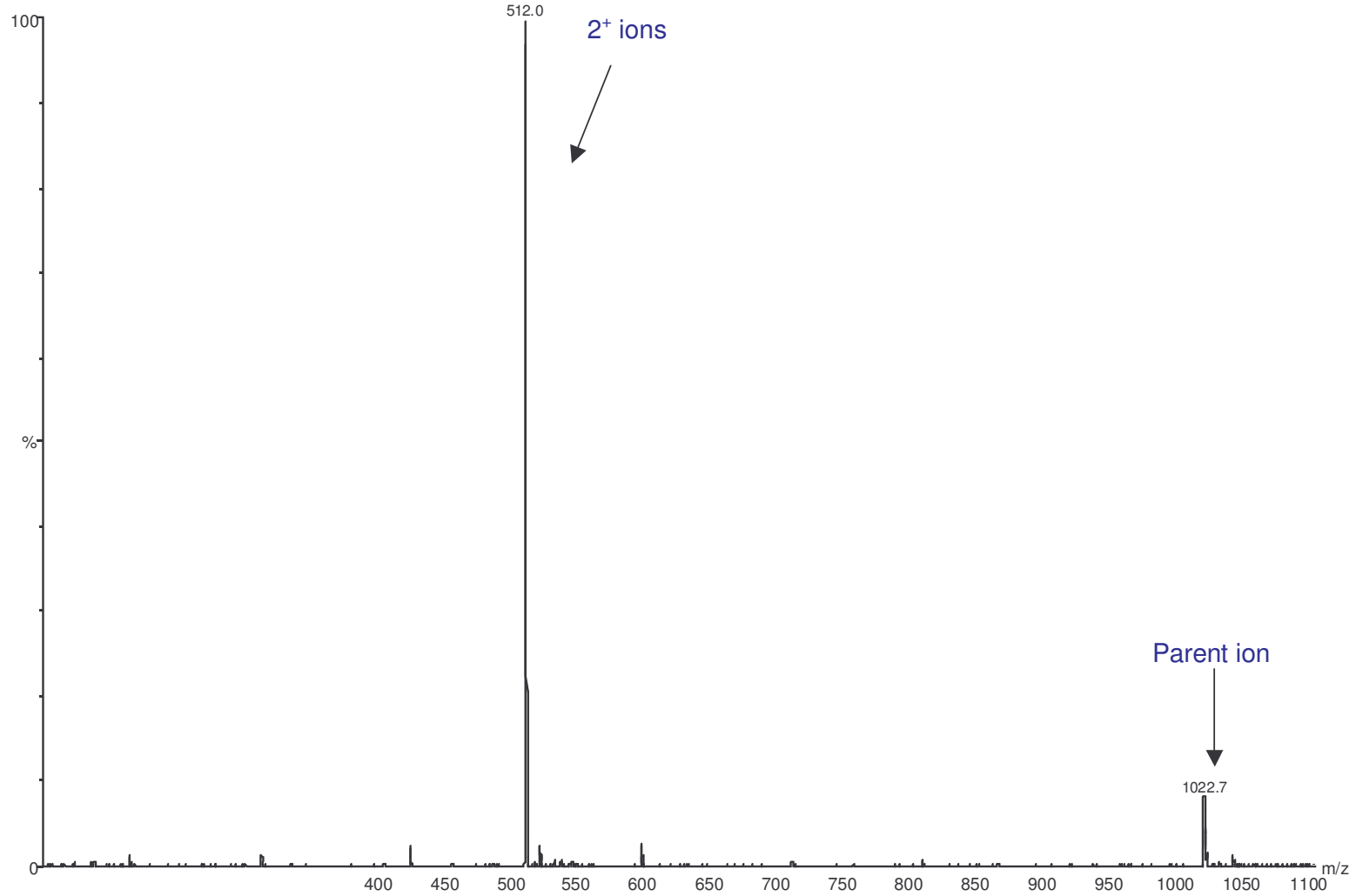


Figure 4.3 LC-MS spectrum of HTLV-I protease cleavage of VLVLPEAKGP (Y114V) peptide.

The data from the VLYLPEAKGP (E117V) peptide (Figures 4.4 and 4.5) and the double mutant, VLVLPVAK, (Figure 4.6) exhibit cleavage products. A very low level of cleavage is observed between the P3 leucine and P2 tyrosine in the E117V mutant with the presence of the peak at 844.3 Da (844.0 predicted) corresponding to the NH_2-YLPVAK-OH cleavage product (Figures 4.4 and 4.5). The double mutant shows a more significant level of cleavage, and is cut between the P3 leucine and P2 valine as indicated by the presence of the peak at 626.6 Da (625.8 predicted) corresponding to the NH_2-VLPVAK cleavage product (Figure 4.6). However, neither substrate was cut at the normal P1-P1' scissile junction between leucine and proline.

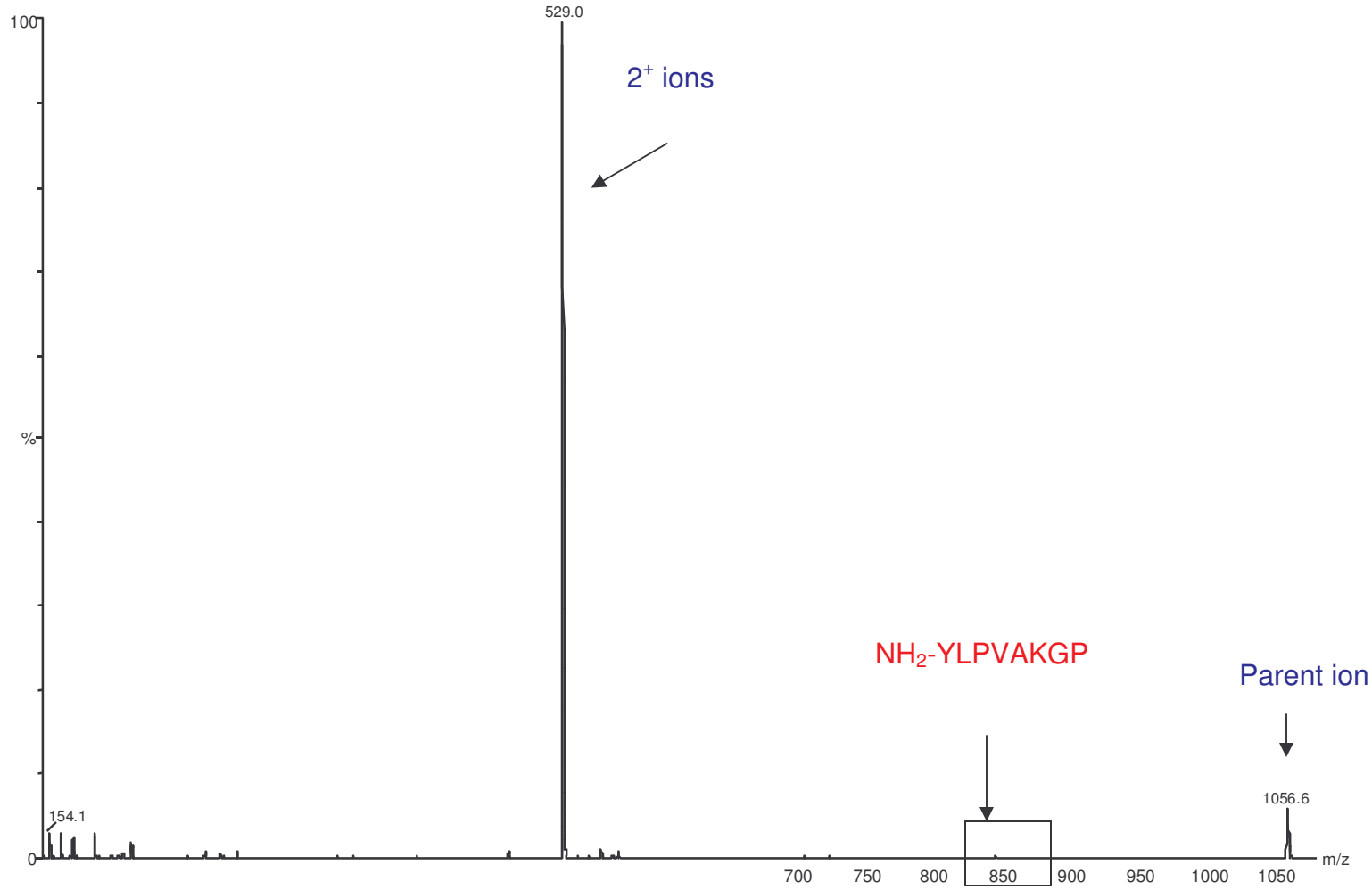


Figure 4.4 LC-MS spectrum of HTLV-I protease cleavage of VLYLPVAKGP (E117V) peptide cleavage.

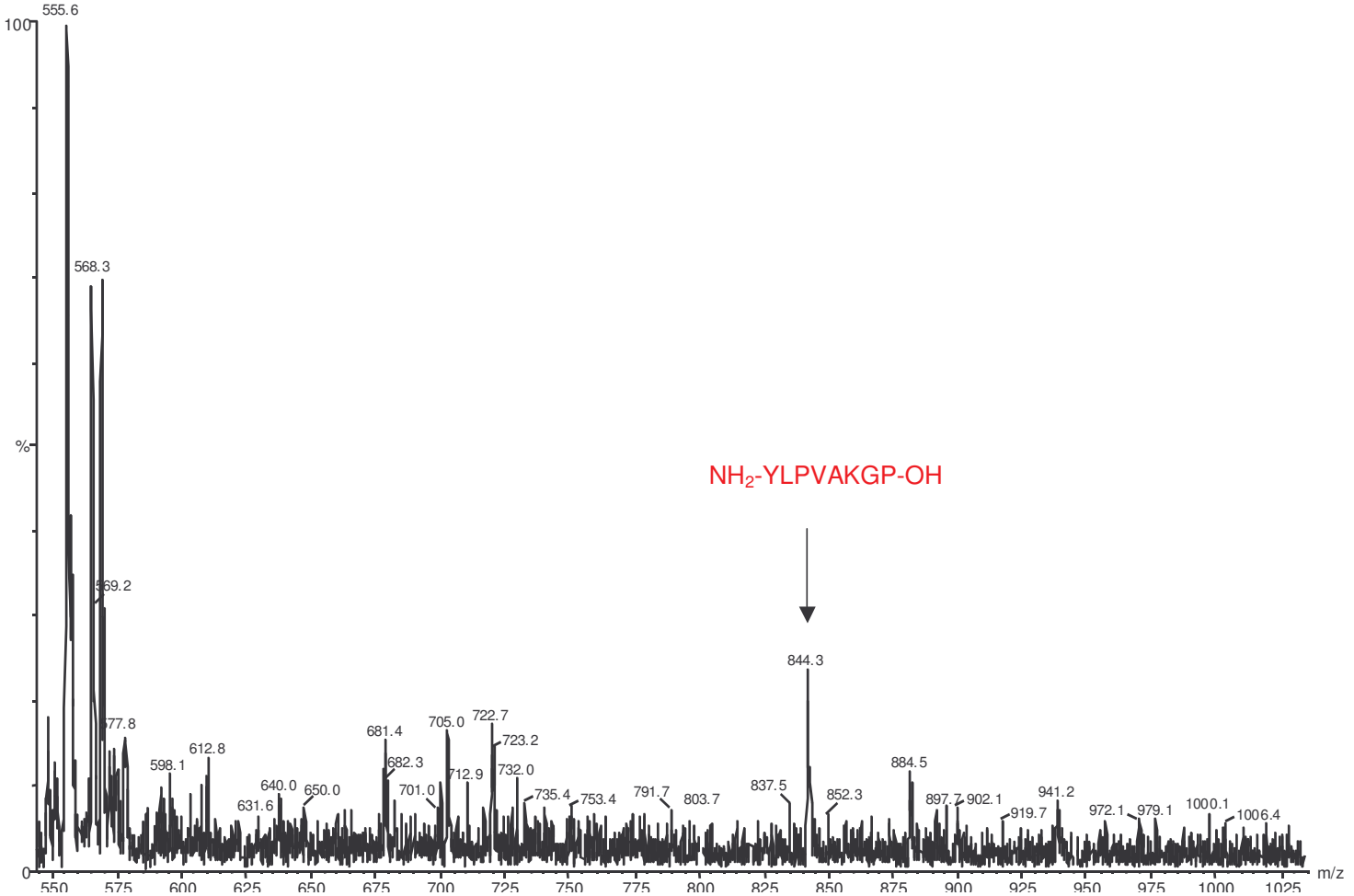


Figure 4.5 Enlarged LC-MS spectrum of HTLV-I protease cleavage of VLYLPVAKGP (E117V) peptide cleavage.

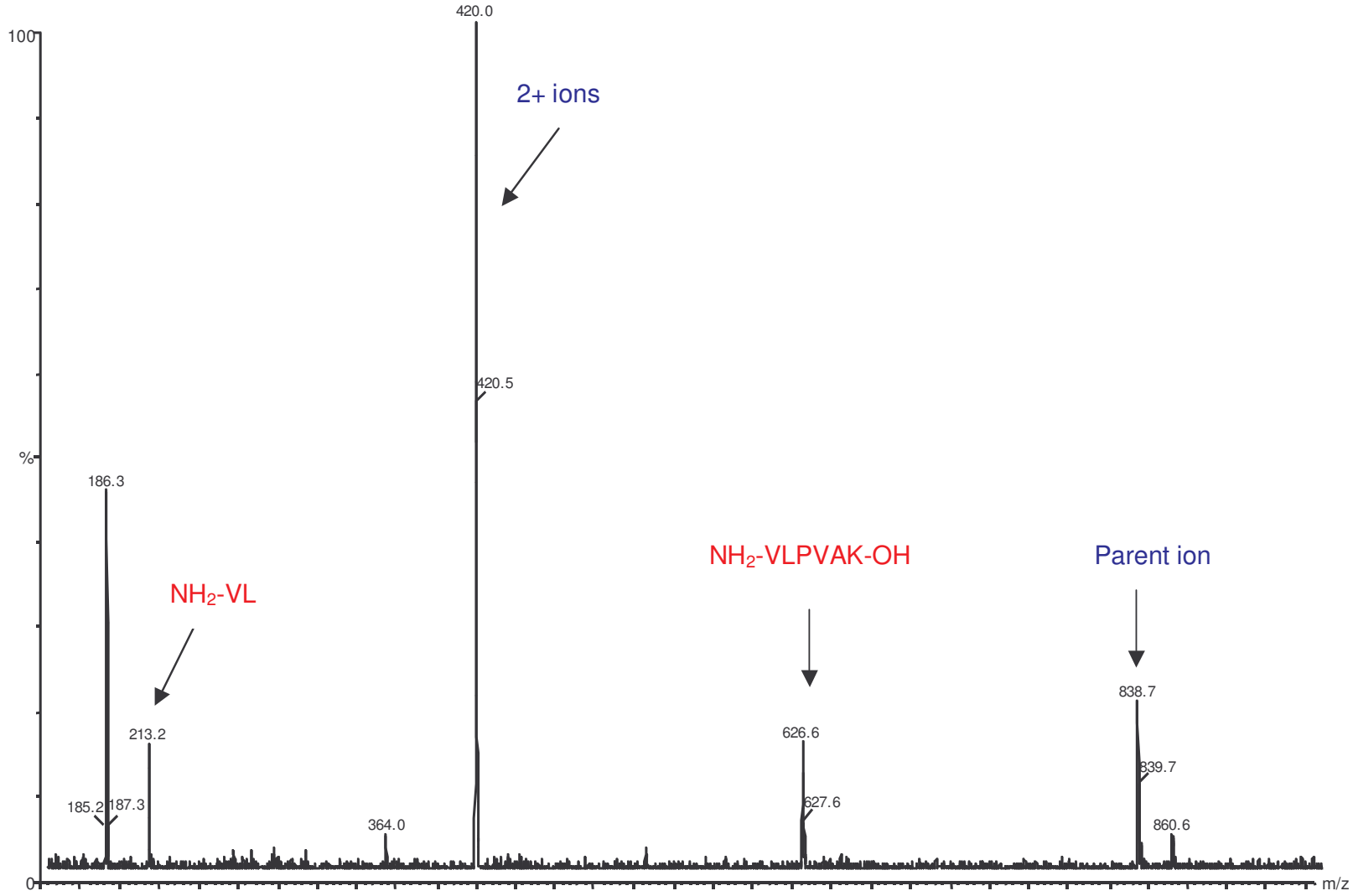


Figure 4.6 LC-MS spectrum of HTLV-I protease cleavage of VLVLPVAK (double mutant) peptide.

To rule out the size of the substrates as a contributing factor to the unusual cleavage pattern, we prepared an extended double mutant:

Ac-QGVLVLPVAK-HMBA-❍ (❍ = peptide linked to bead). This peptide was incubated with HTLV-I protease while attached to the bead and then analyzed by LC-MS. The results in Figure 4.7 show a peak at 456.3 Da corresponding to Ac-QGVL-OH (calculated mass: 457.4). Once again, the substrate was not cut between the P1 leucine and P1' proline, but rather between the P3 leucine and P2 valine.

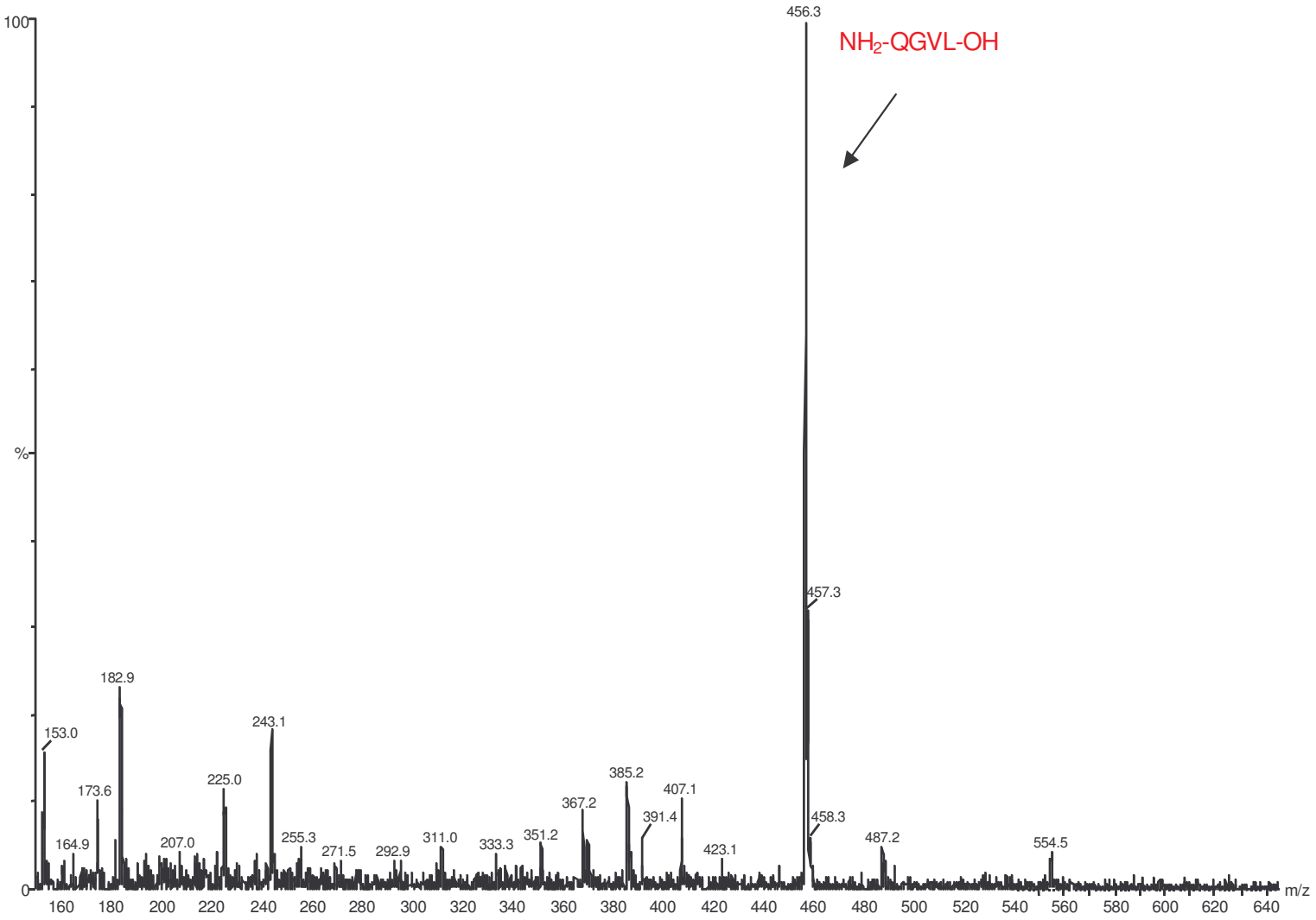


Figure 4.7 LC-MS spectrum of HTLV-I protease cleavage of Ac-QGVLVLPVAK-HMBA-❍ (double mutant) peptide.

Two site-directed mutants of HTLV-I protease (Y114V and E117V) were prepared by Bryan Herger from the L40I mutant to confirm the results obtained from the synthetic peptides. The two new mutants along with L40I, L40I Δ10, and wild-type (wt) protease were analyzed by SDS-PAGE (Figure 4.8). The E117V mutant (Lane 5) has an additional band below the intact protein band that is not present in the wt (Lane 3) or Y114V (Lane 4), indicating that E117V has an additional shorter fragment than the wild-type or Y114V proteases. Lane 6 is a mixture of 50% E117V and 50% Y114V and shows the same band as in the E117V. Lane 1 is the deletion mutant, L40I Δ10. This band should be approximately equal in size to a mutant cleaved at position 114 or 117. Lane 2 is L40I with no autoprocessing. This demonstrates that the C-terminal "tail" domain has undergone autoproteolysis and that it is the Glu^{117} that hinders cleavage of the C-terminal tail. This is in agreement with the synthetic peptide analysis.

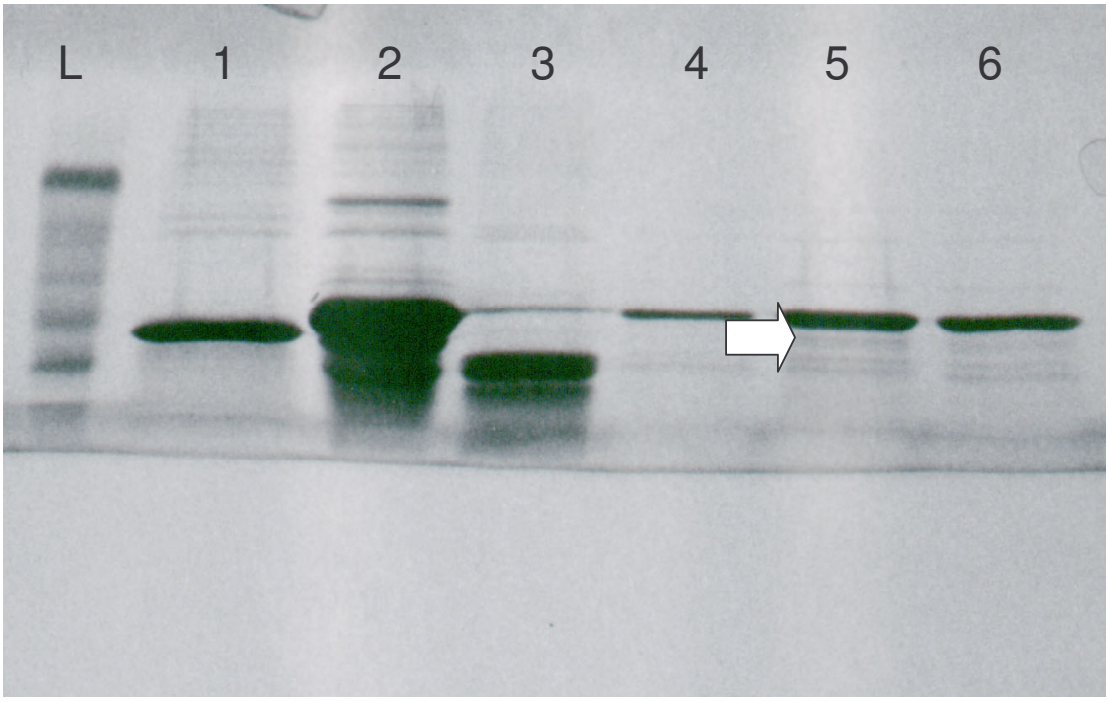


Figure 4.8 Electrophoresis Gel showing bands from HTLV-I wt and mutants (Lane 1: L40I Δ10; Lane 2: L40I; Lane 3: wt; Lane 4: Y114V; Lane 5: E117V; and Lane 6: mixture of Y114V and E117V; L = DNA marker).

To verify the SDS-PAGE result, the mutant cleavage products were analyzed by MALDI (Figures 4.9 and 4.10). The full-length His-PR sequence was at the expected mass (*m/z* =15644.8 for the E117V mutant and *m/z* = 15637.9 for the Y114V mutant). For the E117V mutant (Figure 4.9), an additional peak corresponding to the expected mass of the C-terminal cleavage sequence (*m/z* = 14051.8) demonstrated that cleavage did occur. However, this cleavage is not at the Leu^{115}-Pro^{116} scissile site. Instead, it is shifted two amino acids upstream, which is in agreement with our synthetic peptide mutants. There is no C-tail cleavage for the Y114V mutant (Figure 4.10). However, an

interesting cleavage was observed at around 10 kDa. This peak appears to be another processing site at Leu^{78}-Pro^{79}. The Leu^{72}-Pro^{73} has no cleavage because of the presence a serine residue in the P2 site, which is not tolerated for cleavage. MALDI spectra confirm the autoprocessing site at Leu^{78}-Pro^{79}, where the N-terminal sequence corresponds to the expected mass (*m/z* = 10277.3) in both the E117V and Y114V mutants and the C-terminal processed sequences (*m/z* = 5422.7 for the E117V).

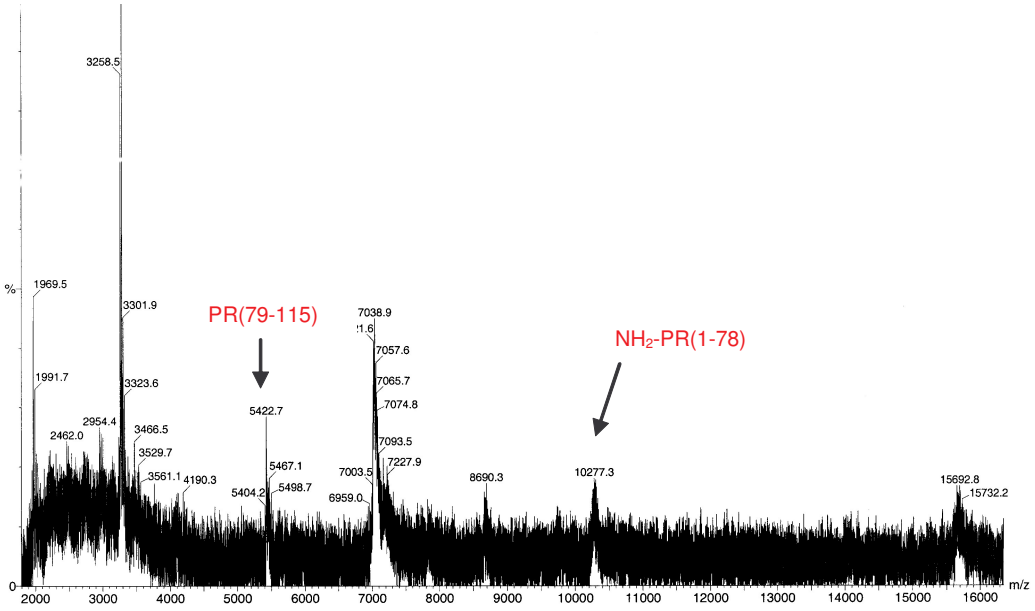


Figure 4.9 MALDI spectrum of E117V. The a peak at *m/z* = 10277 Da (Expected 10244 Da) corresponding to the N-terminal PR(1-78) Leu^{78}-Pro^{79} cleavage fragment and the C-terminal PR(79-115) cleavage fragment at *m/z* = 5422.7 Da.

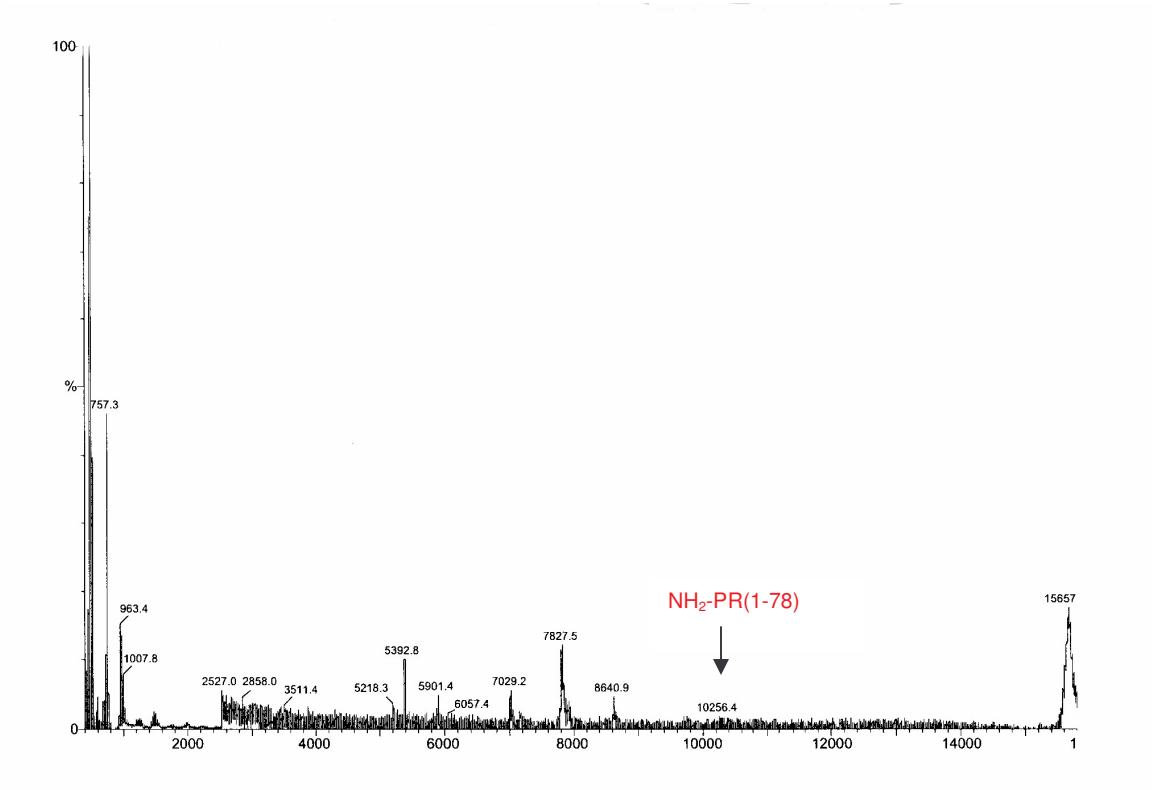


Figure 4.10 MALDI spectrum of Y1147V. The peak at *m/z* = 10256.4 Da (Expected 10244 Da) corresponding to the N-terminal PR(1-78) Leu^{78}-Pro^{79} cleavage fragment and the C-terminal PR(79-115) cleavage fragment at *m/z* = 5422.7 Da.

4.3 Conclusion

As previously noted, Louis et al. concluded that steric considerations play a significant role in HTLV-I protease's affinity for substrate by demonstrating prevention of normal autoproteolysis in a L40I mutant. Isoleucine differs from leucine only in the placement of the methyl branch on the side chain. Further, the list of known substrates

shows a clear preference for a small, aliphatic residue in the P2 position. The peptide having a single mutantion of tyrosine to valine did not show any cleavage products, indicating that the tyrosine alone does not inhibit protease cleavage. The glutamic acid to valine mutation showed only a slight cleavage between the VL and YLPVAKGP, which is not at the typically favored L/P scissile junction. Having a tyrosine to valine mutation along with a glutamic acid to valine mutation in the double mutant leads to a much more prominent cleavage site, showing a more favorable proteolysis to occur at the next site in the nonfunctional TFP peptide domain. The double mutant shows this different cleavage junction between Val^{114} and Leu^{115}. This suggests that charged species in the P2' site have electronic interactions that inhibit proteolytic cleavage and may play a role in the positioning the substrate into the active site, which may explain why the substrate is not cleaved at the normally favored L/P junction.

Structure determination of HTLV-I protease has been an important goal toward understanding the enzymology of the protease. However, structure determination by NMR and x-ray crystallography has been unsuccessful. The L40I mutation that prevents autoprocessing at this site was hoped to have been a key toward forming protease dimers at concentrations large enough for NMR determination. It appears from the MALDI spectra reported here that an additional mutation, L78I, may provide the necessary stability for sufficient dimerization. Further cleavage studies need to be performed to better understand HTLV-I protease behavior.

REFERENCES

Hayakawa, T., Misumi, Y., Kobayashi, M., Yamamoto, Y. & Fujisawa, Y. 1992. Requirement of N- and C-terminal regions for enzymatic activity of human T-cell leukemia virus type I protease. *Eur J Biochem*, **206,** 919-25.

Heidecker, G., Hill, S., Lloyd, P. A. & Derse, D. 2002. A novel protease processing site in the transframe protein of human t-cell leukemia virus type 1 $PR76^{gag\text{-}pro}$ defines the N terminus of RT. *J. Virol.*, **76,** 13101-13105.

Herger, B. E., Mariani, V. M., Dennison, K. J., Shuker, S. B. 2004. The Ten C-Terminal Residues of HTLV-I Protease Are Not Necessary for Enzymatic Activity. *Biochemical and Biophysical Research Communications*.

Kobayashi, M., Ohi, Y., Asano, T., Hayakawa, T., Kato, K., Kakinuma, A. & Hatanaka, M. 1991. Purification and characterization of human T-cell leukemia virus type I protease produced in Escherichia coli. *FEBS Lett*, **293,** 106-10.

Louis, J. M., Oroszlan, S. & Tözsér, J. 1999. Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type I proteinase. *J. Biol. Chem.*, **274,** 6660-6666.

Mariani, V. M. & Shuker, S. B. 2003. Identification of the RT-RH/IN cleavage site of HTLV-I. *Bioche. Biophys. Res. Commun.*, **300,** 268-270.

Précigoux, G., Geoffre, S., Leonard, R., Llido, S., Dautant, A., Destaintot, B. I., Picard, P., Menard, A., Guillemain, B. & Hospital, M. 1993. Modelling, synthesis and biological activity of a BLV proteinase, made of (only) 116 amino-acids. *FEBS Lett.*, **326,** 237-240.

Shuker, S. B., Mariani, V. M., Herger, B. E. & Dennison, K. J. 2003. Understanding HTLV-I Protease. *Chem. Biol.*, **5,** 373-380.

Tözsér, J., Zahuczky, G., Bagossi, P., Louis, J. M., Copeland, T. D., Oroszlan, S., Harrison, R. W. & Weber, I. T. 2000. Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. *Eur. J. Biochem.*, **267,** 6287-6295.

CHAPTER 5

HTLV PROTEASE INHIBITORS

5.1 Peptide Mimetics

Detailed X-ray crystallography has revealed much regarding the inhibitor binding and catalysis of aspartic acid proteases. Inhibitors that mimic the tetrahedral intermediate have been shown to have a high affinity for binding to the active site. In the protease-inhibitor complex, the catalytic aspartic acids lie near the scissile bond between the P1 and P1' positions and the carbonyl group of the glycine in the Asp-Thr-Gly catalytic triad forms a hydrogen bond with the amino group of the inhibitor at positions P1 and P2' (Figure 5.1) (Swain et al., 1990).

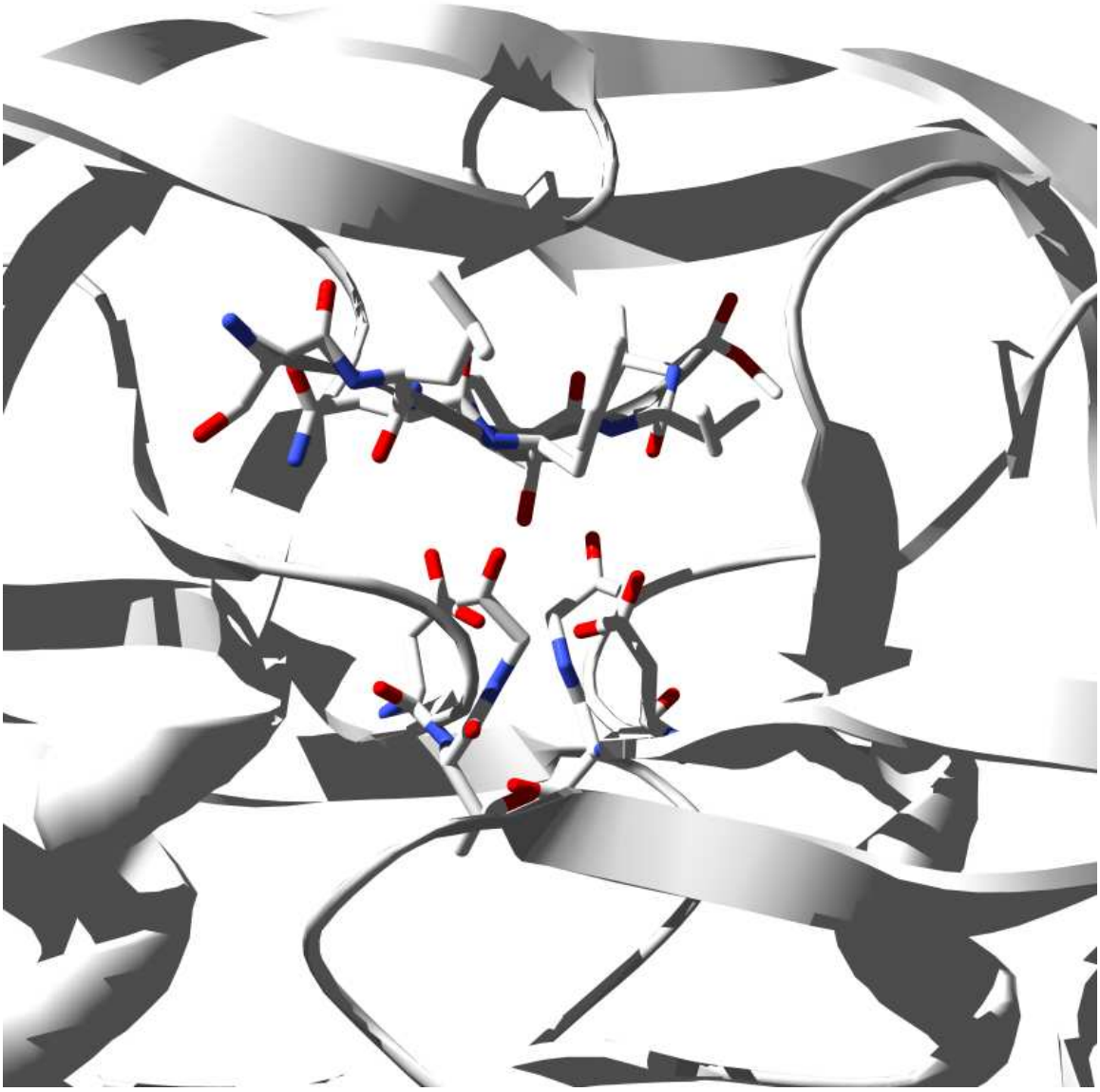
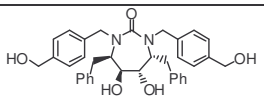
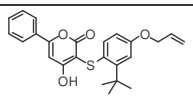
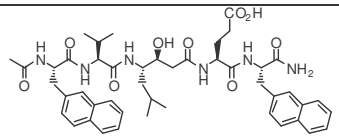
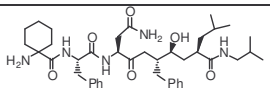
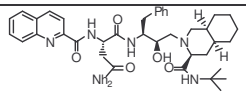


Figure 5.1 HIV-1 protease (PDB 7HVP) with inhibitor JG-365 (shown as sticks; cpk format) bound in the acitive site. The catalytic triads (Asp-Thr-Gly) are also shown as sticks; cpk format.

5.2 Statine-, 4-Amino-3-hydroxy-5-phenylpentanoi acid-, and Hydroxyethylamine-based HTLV-I Inhibitors

Kinetic results for several HIV and pepsin protease inhibitors assayed with HIV and HTLV-I proteases have been reported (Daenke et al., 1994, Louis et al., 1999, Teyura et al., 2002). Table 5.1 gives the K_i data for inhibitors that were assayed with HTLV-I protease. Based on this data, the best inhibitor was found to be the HIV-1 inhibitor: JG-365 (Ac-Ser-Leu-Asn-Phe-CH(OH)CH_2NH-Pro-Ile-Val-OMe) with K_i = 7.2 nM (Ding et al., 1998). However, as reported by Julie Ha, these data may not have been correct due to correction errors in protease concentration used in the assays (Ha et al., 2002).

Table 5.1 K_i values for compounds that have been tested for inhibition of HTLV-I protease.

Inhibitor	Structure	HTLV-I K_i (μM)	HIV-1 K_i (μM)	Ref
DMP 323		>10	0.0003	(Louis et al., 1999)
MES13-099		0.24	0.007	(Teruya et al., 2002)
LP-149		>10	0.0017	(Louis et al., 1999)
UK-88,897		>10	0.012	(Louis et al., 1999)
Ro-31-8959		>10	0.0004	(Louis et al., 1999)
N-1270	Arg-Val-Leu*-Phe-Glu-Ala-Nle-NH_2[b]	127	0.05	(Teruya et al., 2002)
N-1465	Ac-Thr-Ile-Nle*-Nle-Gln-Argb	12.7	0.78	(Teruya et al., 2002)
N-1460	Ser-Gln-Asn-Phe*-Pro-Ile-Val-Gln[b]	7.3	1	(Teruya et al., 2002)
1	Boc-Val-Val-Phe(PC)Phe-Val-Val-NH_2[a]	0.08	0.0004	(Louis et al., 1999)
JG-365	Ac-Ser-Leu-Asn-Phe-CH(OH)CH_2NH-Pro-Ile-Val-OMe	0.006	0.00066	(Ding et al, 1998, Teruya et al., 2002,)
2	Lys-Thr-Lys-Val-Sta-Val-Gln-Pro-Lys	>10	>10	(Louis et al., 1999)
3	Pro-Pro-Cys-Val-Phe-Sta-Ala-Met-Thr-Met	>10	NDc	(Louis et al., 1999)
4	Pro-Tyr-Val-Phe-Sta-Ala-Met-Thr	>10	12.5	(Louis et al., 1999)
5	Ala-Pro-Gln-Val-Sta-Val-Met-His-Pro	0.05	0.13	(Louis et al., 1999)
6	Ac-Pro-Gln-Ile-Thr-Leu-Trp-Gln-Arg-Pro-NH_2	1700	IC_{50} > 2000	(Daenke et al., 1994)
7	Ac-Thr-Leu-Asn-Phe	252	IC_{50} = 1500	(Daenke et al., 1994)
8	Ac-Thr-Val-Ser-Phe-Asn-Phe	217	6.1	(Daenke et al., 1994)
N-1395	Ac-Leu-Val-Phe-H	93	IC_{50} = 0.90	(Teruya et al., 2002)
9	Pro-Val-Ile-Pro-Leu-Asp-Pro-Ala-Arg-Arg-Pro-Val	87.2	17.1	(Daenke et al., 1994)
10	Gln-Met-Gln-Gly-Val-Leu-Tyr-Leu	70	IC_{50} > 400	(Daenke et al., 1994)
11	Ac-Leu-Lys-Ala-Gln-Ile-His-Phe	50.6	no inhibition	(Daenke et al., 1994)
12	Tyr-Leu-Pro-Glu-Ala-Lys-Arg-Pro-Pro-Val-Ile-Leu	5.7	84	(Daenke et al., 1994)

[a] Asterisk denotes a reduced peptide bond, [b] PC denotes replacement of the amide bond with –P(O)(OH)-, and [c] ND = not determined.

In addition to compounds tested in Table 5.1, five other peptidomimetic pepsin inhibitors containing a statine isostere have also been tested against HTLV-I protease (Ding et al., 1998). Table 5.2 shows the inhibition activity of these five inhibitors. The CS-I-25 (Iva-Lys-Val-Sta-Ala-Iaa) inhibitor was found to be a reasonable inhibitor with a K_i = 7.2 ± 0.2 nM (10 nM with PR-His-tag present). However, as with the JG-365, the CS-I-25 was probably improperly corrected for protease concentration (Ha et al. 2002).

Table 5.2 Inhibition of HTLV-I Protease by Pepsin Inhibitors.

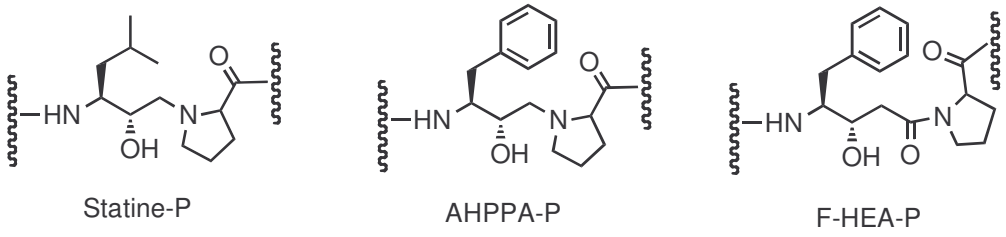
Inhibitor	Structure[a]	K_i (pepsin)[b] (nM)	K_i (HTLV-I PR) (nM)
CS-I-51	Iva-Val-Val-Sta-Lys-Iaa	0.72	466 ± 74
CS-I-22	Iva-Val-Nle-Sta-Ala-Iaa	0.030	138 ± 10
CS-I-25	Iva-Lys-Val-Sta-Ala-Iaa	19.2	7.2 ± 0.2
CS-I-27	Boc-Lys-Val-Val-Sta-Ala-Iaa	0.22	142 ± 6
CS-I-52	Iva-Val-Val-Sta-Ala-Lys-OMe	0.10	7.200 ± 1.500

[a] Iva, isoveryl $(CH_3)_2CHCH_2CO$; Nle, norleucine $CH_3(CH_2)_3CH(NH_2)COOH$; Sta (3S,4S)-statine $(CH_3)_2CHCH_2(NH_2)CH(OH)CH_2COOH$; Iaa, isoamylamine $NH_2CH_2CH_2CH(CH_3)_2$; Boc, isobutyl oxycarboxyl $(CH_3)_3COCO$. [b] (Kuzmic et al., 1991)

Statine, 4-amino-3-hydroxy-5-phenylpentanoic acid (AHPPA), and phenyl-hydroxyethylamine (F-HEA) were expected to be potent inhibitors of HTLV-I protease based on previous work reported by Ding et al. Akaji *et al.* tested several HEA inhibitors

against HTLV-I protease, in which the best inhibitor resulted in a K_i = 38 ± 8 nM (Akaji et al., 2003). The JG-365 has a 10-fold increase in potency over the Akaji *et al.* HEA inhibitor. However, as previously stated this may have been miscalculated (Ha et al. 2002).

Statine, AHPPA, and HEA are nonstandard amino acids that mimic the tetrahedral intermediate of aspartic acid protease catalyzed hydrolysis of the peptide substrate. These structures are illustrated below in Figure 5.2, which shows statine, AHPPA, and F-HEA centerpieces attached to the P1' proline.



Statine-P AHPPA-P F-HEA-P

Figure 5.2 Statine-Pro, AHPPA-Pro, and Phe-HEA-Pro peptide inhibitor centerpieces.

The statine and AHPPA peptides are easily synthesized by solid-phase methods. The statine iso-butyl side chain is expected to bind in the S1 pocket normally occupied by a P1 leucine in the natural substrates. Statine and AHPPA are also expected to have

tight S1 binding. Little is known about the structure-activity relationships (SAR) at each position in HTLV-I protease. A small library of nine peptides were synthesized to investigate the inhibition potency of peptide mimetics, which will lead to the development of potent nonpeptide inhibitors.

The peptides have the general structure:

Ac-X-Y-Sta-Pro-Z-R

Ac-X-Y-AHPPA-Pro-Z-R

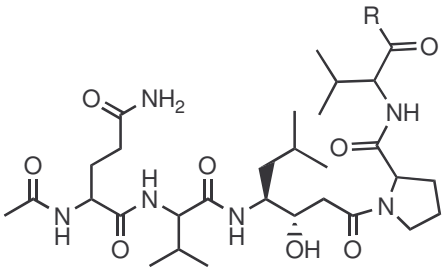
Ac-X-Y-F-HEA-Pro-Z-R

Where Ac = acetyl, X (P3), Y (P2), and Z (P2') are various amino acids with their corresponding position, R = NH_2Et, OMe, or OH. Peptides corresponding to the MA/CA, TF1/PR, and PR/TFP cleavage junctions were prepared (Table 5.3).

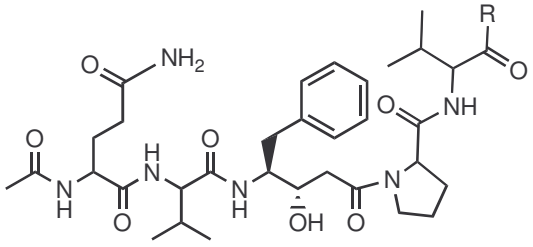
Table 5.3 HTLV-I Cleavage Junctions for Peptide Mimetics.

JUNCTION	CLEAVAGE SITE
MA/CA	APQVL/PVMHP
TF1/PR	PASIL/PVIPL
PR/TFP	PPVIL/PIQAP

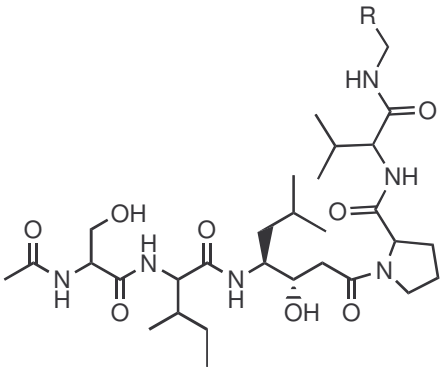
The X (P3) contains Qln, Ser, or Val; the Y(P2) and Z(P2') consists of Val or Ile; all P1' are Pro; and the P1 site have statine, AHPPA, or F-HEA. The statine- and AHPPA-based peptide structures are shown in Figure 5.3.



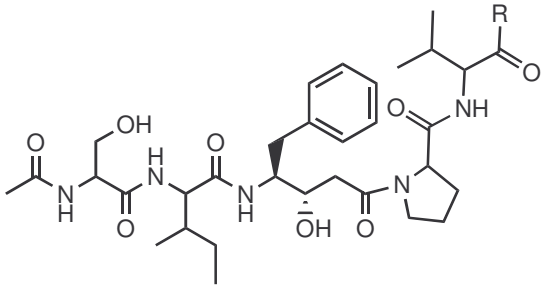
1 **QV-Sta-PV**



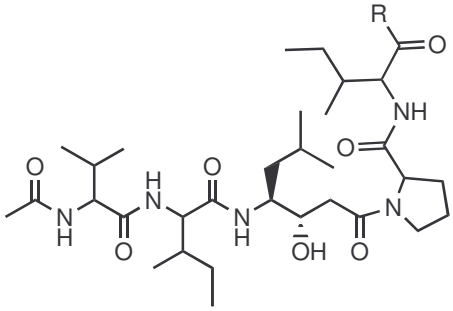
4 **QV-AHPPA-PV**



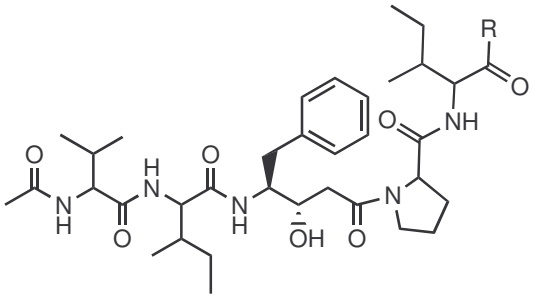
2 **SI-Sta-PV**



5 **SI-AHPPA-PV**



3 **VI-Sta-PI**



6 **VI-AHPPA-PI**

Figure 5.3 Structures of statine and AHPPA inhibitors corresponding to MA/CA, TF1/PR, and PR/TFP cleavage junctions. R = NH_2Et, OMe, or OH.

5.3 Solid-Phase Synthesis of Statine and 4-Amino-3-hydroxy-5-phenylpentanoic acid-based Inhibitors

The statine- and AHPPA-based peptides were first synthesized on solid support using Argopore amine-substituted beads (Wellings & Atherton, 1997). To each bead, 4-hydroxymethylbenzoic acid (HMBA) was attached as a cleavable linker by base-catalysis with diisopropylethylamine (DIPEA). Then the first 9-fluorenylmethoxycarbonyl (Fmoc)-protected amino acid was attached to the linker by forming the symmetrical anhydride with N,N'-dicyclohexylcarbodiamide (DCC) and adding this to the HMBA-linked bead with 4-dimethylaminopyridine (DMAP). The Fmoc was deprotected with 20% piperdine (mechanism shown in Figure 5.4). Following deprotection, the next amino acid was added by base catalysis with DIPEA and a coupling agent---2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) for primary amino acids and 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) for subsequent amino acids.

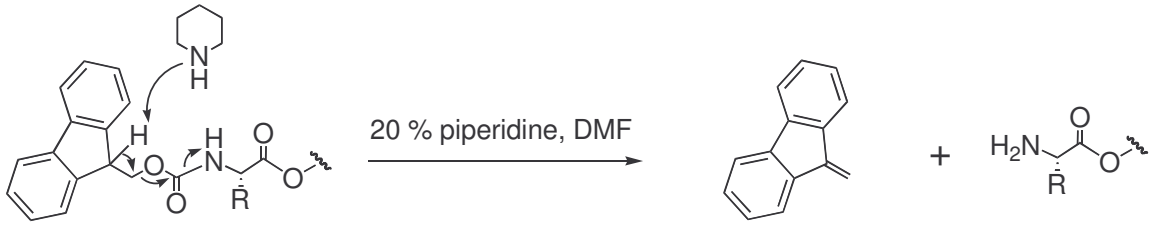


Figure 5.4 Mechanism for Fmoc-deprotection.

The iterative process of addition and deprotection was repeated until the final desired amino acid sequence was completed. This process is shown in the general scheme illustrated in Figure 5.5.

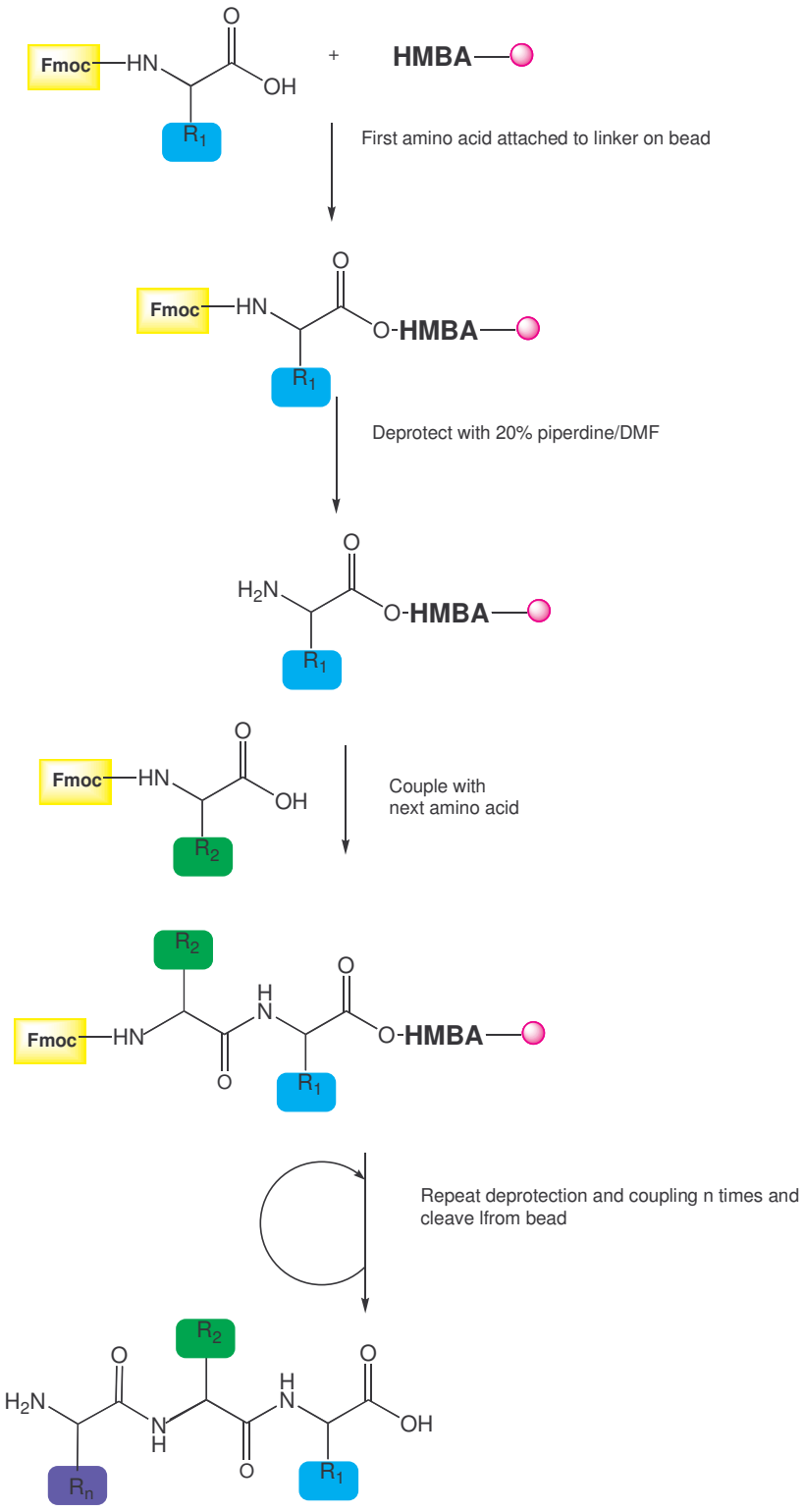
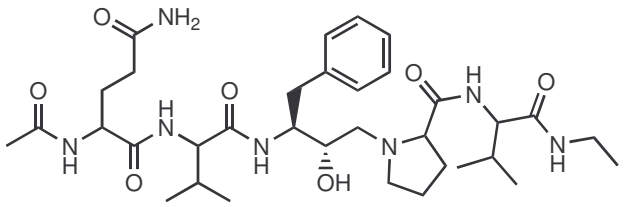


Figure 5.5 General scheme of solid-phase synthesis.

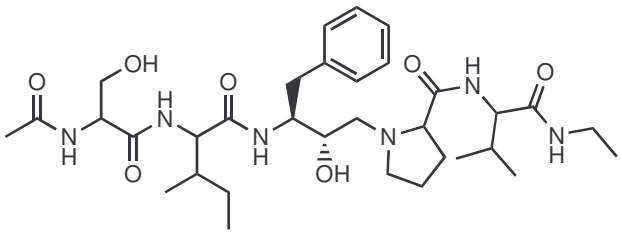
The yields were extremely poor so another route was attempted. In the second round of peptide synthesis, pre-loaded Val(Ile)-Tenta-Gel beads were used instead of the amine substituted Argopore beads. The only deviations from the synthesis from the previous method were that equal molar amounts of HOBt and TBTU were used in each addition step and TFA was used to cleave the peptide from the bead. This gave much better yields. However, only a few of the peptides had yields large enough to purify by HPLC.

5.4 Phenylhydroxyethylamine Synthesis

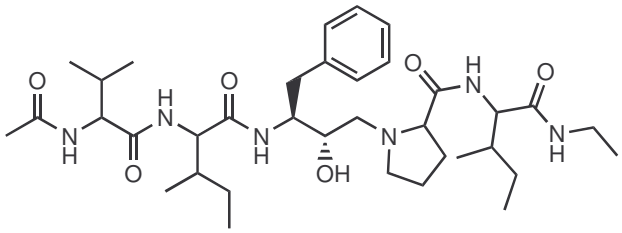
Similar to the statine- and AHPPA-based inhibitors, the F-HEA-based inhibitors also mimic the MA/CA, TF1/PR, and PR/TFP cleavage junctions. The structures for these three inhibitors are shown in Figure 5.6.



7 QV-F-HEA-PV



8 SI-F-HEA-PV



9 VI-F-HEA-PI

Figure 5.6 Structures of Phe-HEA inhibitors corresponding to MA/CA, TF1/PR, and PR/TFP cleavage junctions.

The HEA synthesis followed a two-part strategy (Figure 5.7). The first step followed standard Fmoc solid-phase chemistry on an amine-substituted Argopore bead with the cleavable HMBA linker. The second step involved attaching N-Tosyl-L-phenylalanine chloromethyl ketone (TPCK), which contains the starting block for Phe-HEA. As shown in scheme in Figure 5.8, the ketone of the TPCK was first protected as an acetal and then attached to either Pro-Val-HMBA-❍ or Pro-Ile-HMBA-❍ following a similar strategy by Daniel Rich *et al.* (Rich et al., 1992). Next, the Tosyl protecting group was removed by an electron transfer reaction with $NaBH_4$, $h\nu$, and 2,7-dimethoxynapthalene as an electron donor (Hamada et al., 1986). Once it was deprotected, the final two amino acids were added to the HEA centerpiece via Fmoc synthesis.

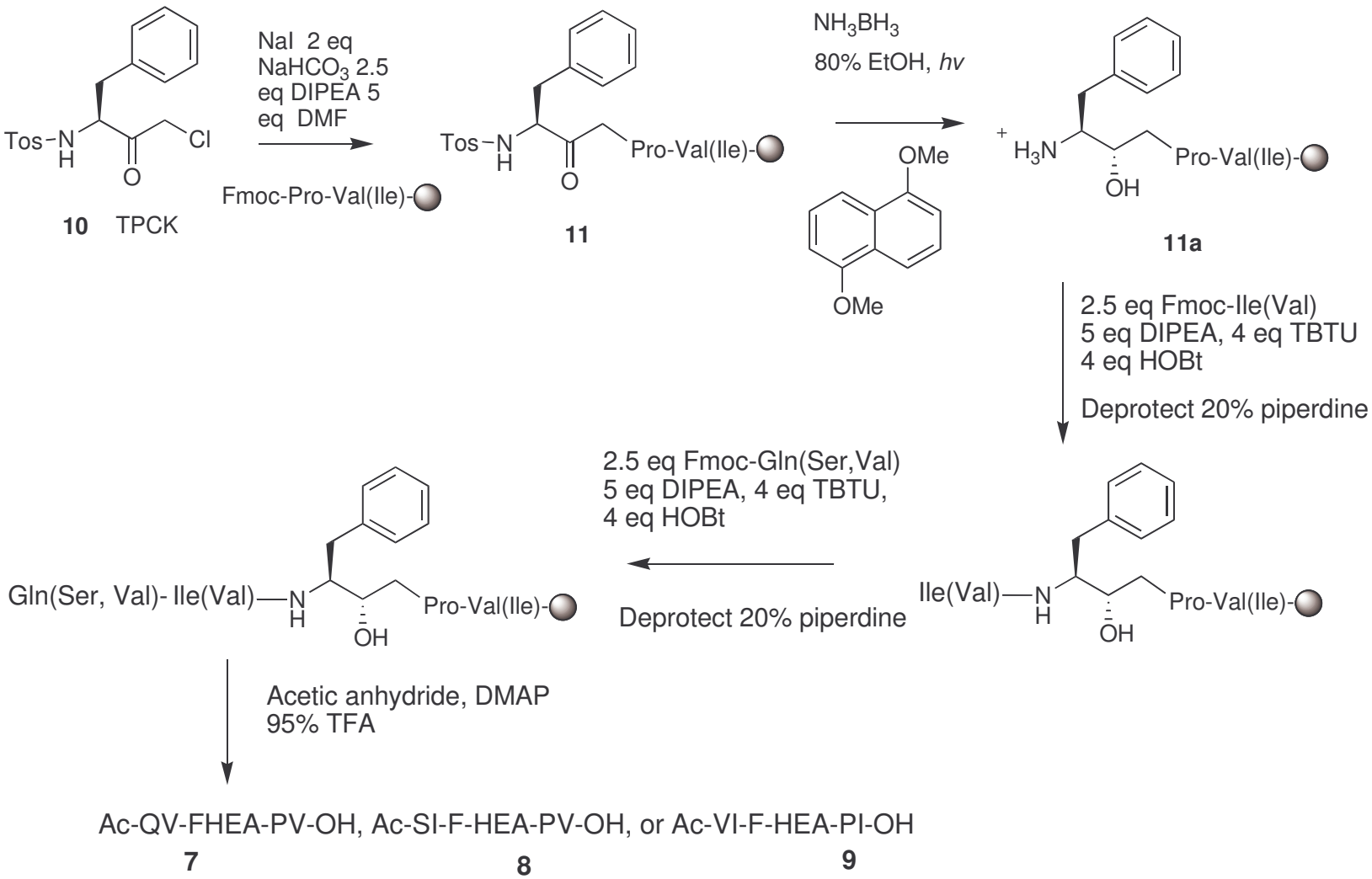


Figure 5.7 Scheme for HEA inhibitors.

Before the TPCK could be added to the beads containing either Pro-Val or Pro-Ile, the reaction status was checked. TPCK (**10**) was added to proline following a similar procedure outlined in the scheme in Figure 5.8 to synthesize compound **11**. The reaction shown in Figure 5.8 was completed with 2 equivalents of proline rather than reacted with the proline attached to the bead (**11**) as shown in Figure 5.8.

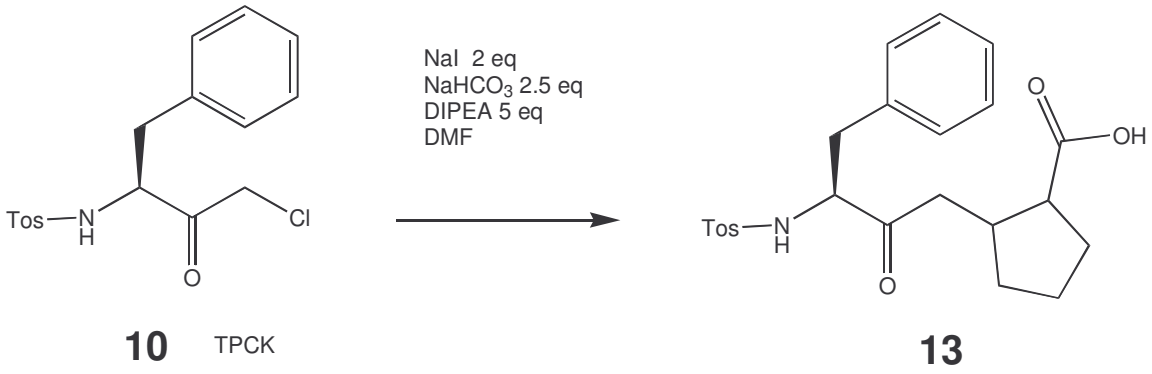


Figure 5.8 Reaction of TPCK (**10**) to TPCK-Pro (**13**).

Comfirmatiion of compound **13** was verified by ^{1}H NMR (see experimental section Chapter 9). There was an upfield shift in the methylene protons of **10** from δ 4.25 to overlapping the β protons at δ 3.42. In addition the presence of the praline hydroxyl proton was seen at δ 10.60.

5.5 Fluorescence Assays of Synthetic Peptide Inhibitors

Following cleavage with ethylamine, the statine and AHPPA peptides were washed and purified by HPLC. The purity was confirmed by LC/MS. Inhibition by the synthesized peptide mimetics was assayed by fluorescence resonance energy transfer (FRET) with a CA/NC fluorogenic substrate: 4-(4-Dimethylaminophenylazo)benzoyl-γ-aminobutyric acid-PQVLPVMH-5-(2-aminoethylamino)-1-napthalenesulfonic acid (Dabcyl-GABA-PQVLPVMH-EDANS). On this substrate, EDANS serves as the fluorophore and is quenched by Dabcyl. This substrate is shown in Figure 5.9.

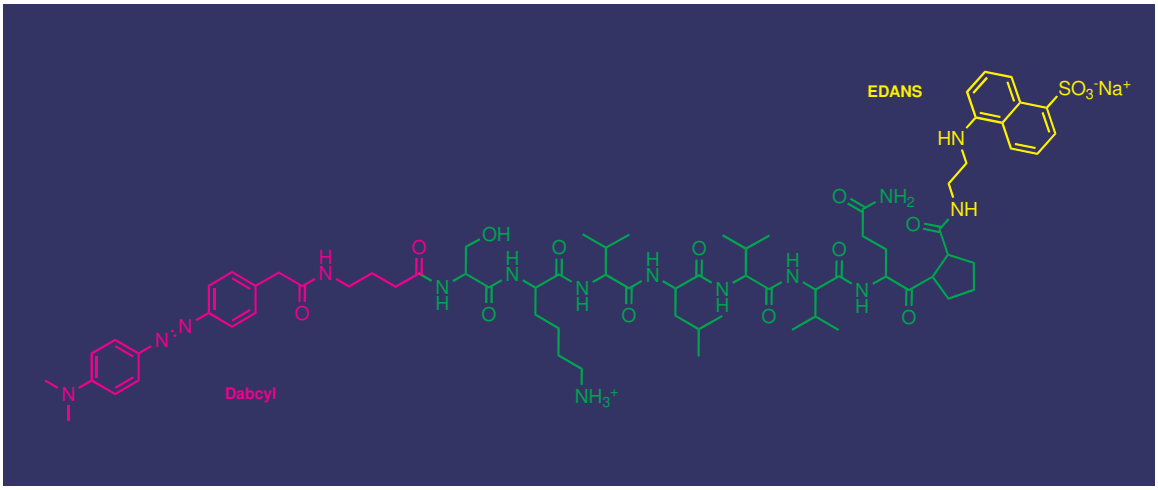


Figure 5.9 Fluorogenic substrate (Dabcyl-GABA-TKVLVVQP-EDANS) for HTLV-I protease. This substrate is based on the CA/NC junction, TKVLVVQP (green), with fluorophore EDANS in yellow, and quencher Dabcyl in violet.

When the peptide is cleaved, EDANS is no longer quenched by Dabcyl. The inhibition constants of each inhibitor can then be measured by monitoring the amount of fluorescence present by the substrate. One of the inhibitors with the highest percent inhibition (QV-Sta-PV) and QV-AHPPA were taken to the CDC for *in vitro* analysis on HTLV-I infected cell lines.

5.6 Preliminary Fluorescence Results

The FRET assay was used to determine the percent of inhibition of the statine and AHPPA inhibitors. In this assay, the inhibition of crude unpurified peptides were measured against HTLV-I substrate. The percent inhibition was based on native substrate cleavage set to 0%. The relative inhibition of the peptides were then measured against the substrate. Table 5.4 shows the results of this assay.

Table 5.4 Results of Statine, AHPPA, and HEA Inhibition Assays

Corresponding Cleavage Junction	Inhibitor	Percent Inhibition
	No inhibitor present	0%
MA/CA	Ac-Q-V-Sta-PV	87.6%
TF1/PR	Ac-S-I-Sta-PV	63.1%
PR/TFP	Ac-V-I-Sta-PI	34.5%
MA/CA	Ac-Q-V-AHPPA-PV	61.6%
TF1/PR	Ac-S-I-AHPPA-PV	41.9%
PR/TFP	Ac-V-I-AHPPA-PI	42.7%
MA/CA	Ac-QV-F-HEA-PV	46.5%
TF1/PR	Ac-SI-F-HEA-PV	31.5%
PR/TFP	Ac-VI-F-HEA-PI	67.3%
CS-I-25	Iva-KV-Sta-A-Iaa	86.4%

These data indicate that the statine MA/CA inhibitor had the best competitive inhibition with a result of only 12% cleavage of the native substrate. This is comparable to the CS-I-25 pepsin inhibitor, which showed 14% cleavage of the substrate. The TF1/PR HEA inhibitor was the least effective showing 68% of the native substrate cleaved.

5.7 Inhibition Constants of Purified Peptides

The K_i values for purified inhibitors were determined by measuring inhibitor activity at concentrations in the range of 1 to 5 μM against varying substrate concentrations of 5 to 20 μM. Measurements were taken by monitoring the fluorescence of the Dabcyl-GABA-PQVLPVMH-EDANS substrate. The K_i values were then determined by a method described by Dixon (Segel, 1975). The Ki was also determined for CS-I-25. Results are shown in Figures 5.10 through 5.12 and Table 5.5.

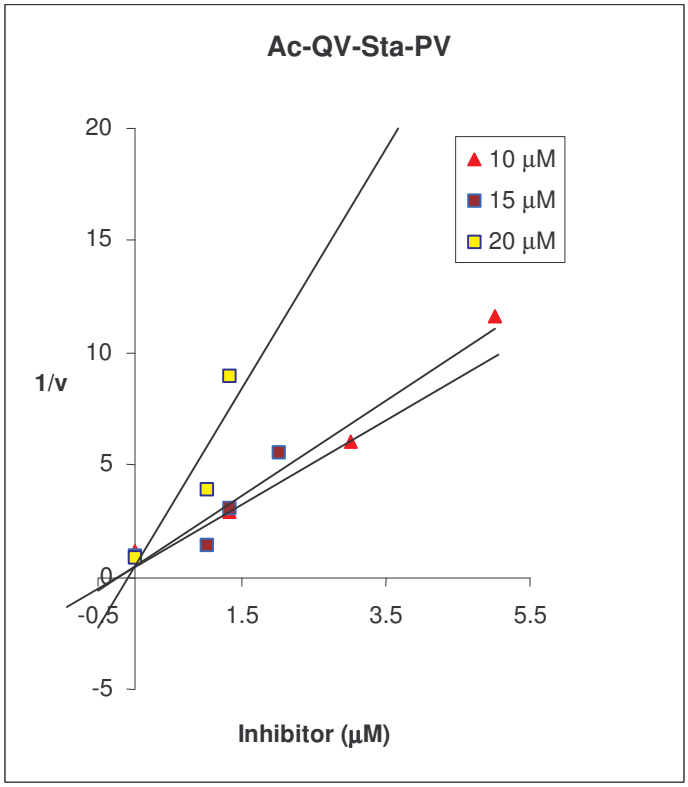


Figure 5.10 Dixon plot for Ac-QV-Sta-PV inhibitor.

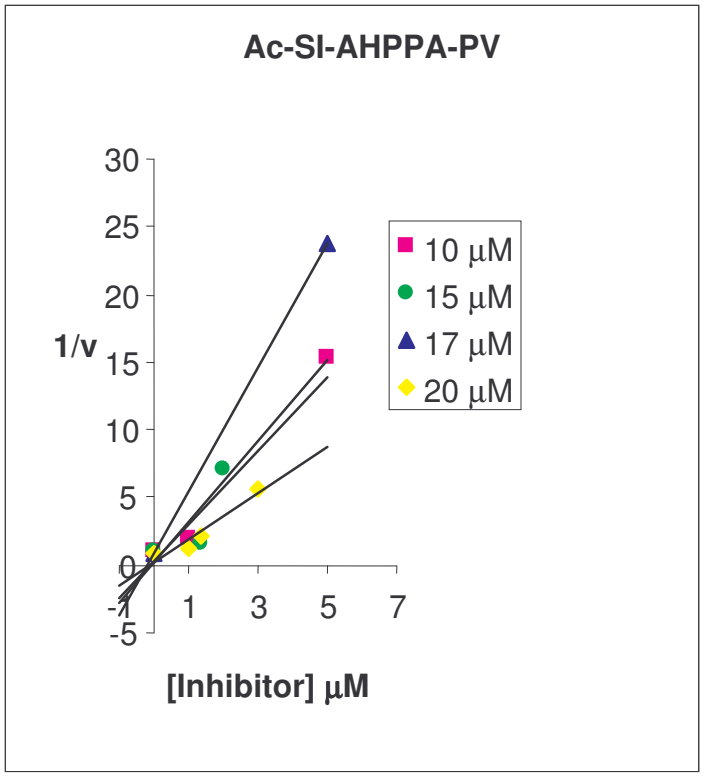


Figure 5.11 Dixon plot for Ac-SI-AHPPA-PV inhibitor.

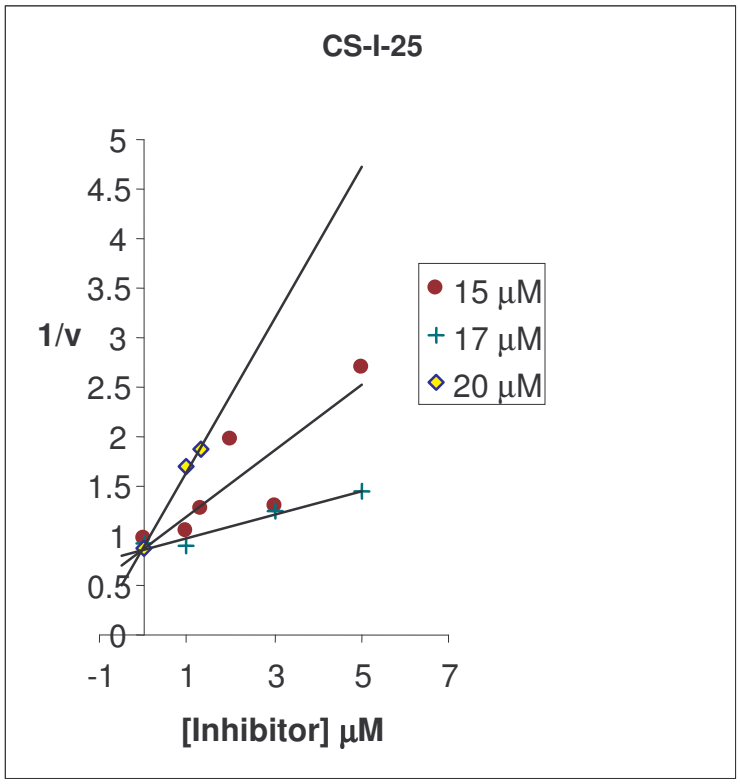


Figure 5.12 Dixon plot for CS-I-25 pepsin inhibitor.

Table 5.5 K_i's of the Inhibitors of HTLV-I Protease.

Inhibitor	**K_i (His-Pr) (nM)**
Ac-QV-Sta-PV	29 ± 4
Ac-SI-AHPPA-PV	38 ± 6
CS-I-25	30 ± 3

As can be seen in Table 5.5 the MA/CA Ac-QV-Sta-PV (K_i = 29 ± 4 nM) has the same inhibition constant as the CS-I-25 pepsin inhibitor (K_i = 30 ± 3 nM). The previously reported 10 nM inhibition for the CS-I-25 by Ding et al. is on the same order of magnitude as the 30 nM reported in this work. In addition, Louis et al., also assayed a HTLV-I protease MA/CA statine-based inhibitor (APQV-Sta-PVMHP). This peptide (K_i = 50 nM) had both the P1 Leu and P1' Pro replaced with statine while the Shuker MA/CA inhibitor only had the P1 Leu replaced with statine. The Shuker inhibitor yielded slightly better inhibition. The similarity in inhibition of the MA/CA statine inhibitor and CS-I-25 was also shown with the percent inhibition assay shown in Table 5.4. Akaji et al described synthesis of a hydroxyethylamine isostere that gave a K_i = 38 ± 8 nM when assayed against HTLV-I protease (Akaji, 2003).

It was expected that the inhibitors synthesized would show K_i's of less than 1 nM based on work by Rich et al. on HIV protease inhibitors. Unfortunately, none of the purified peptides showed this level of inhibition.

5.8 Inhibition of HTLV-I Protease in Cell Culture

Synthesized peptides with good *in vitro* activity were tested in two different cell-based assays. It has been previously shown that HTLV-I can be passed from irradiated virus-producing cell lines such as MT-2 and HUT-102 by co-culture with uninfected peripheral blood mononuclear cells (PBMCs) (Kitamura et al., 1993; Lal & Heniene, 1996). This assay was used to determine if the inhibitors prevent the spread of HTLV-I. The effectiveness of the protease inhibitors was ascertained by monitoring culture supernatants for the production of cell free viral antigen using a commercial p19 ELISA (Zeptometrix).

5.9 Results of Inhibition of HTLV-I Protease in Cell Culture

The PBMC co-cultured with irradiated HUT-102 HTLV-I infected cell lines were treated biweekly with 5, 10, and 20 μM concentrations of the Ac-QV-Sta-PV, CS-I-25, and a nonbonding peptide, Leu-Leu-Leu (L_3). In addition, a HUT-102 cell line (no co-culture) was also treated biweekly with the same concentrations of inhibitors. The co-

cultured irradiated HUT-102s and PBMCs were assayed by p19 antigen ELISA. The results of the ELISA are shown in Figure 5.14.

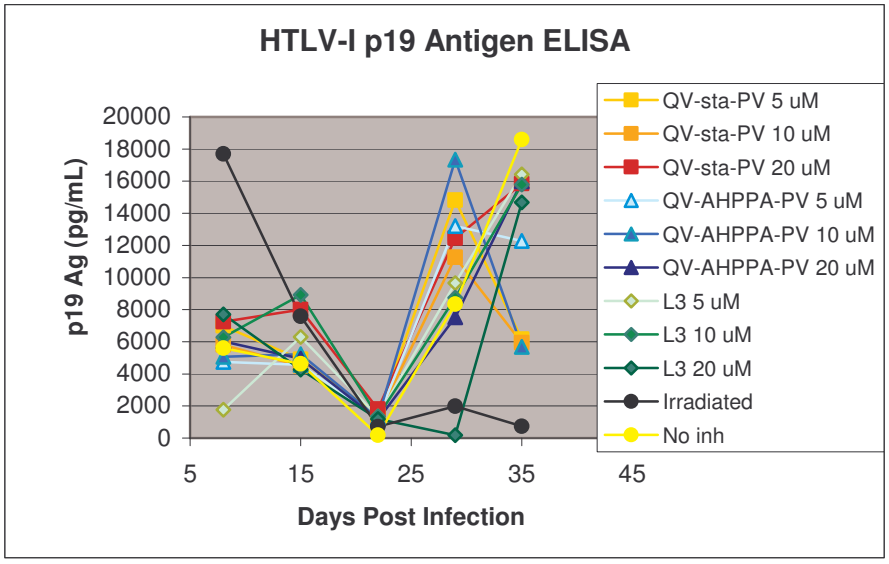


Figure 5.14 Results of p19 ELISA following five weeks of treatment.

As can be seen in Figure 5.14, this five-week treatment is beginning to show signs of inhibition of protease inhibitors in an *in vitro* cell culture assay. The QV-Sta-PV and QV-AHPPA-PV show inhibition at 5 and 10 μM compared to the control peptide L_3 (5, 10, and 20 μM). The L_3 controls are all show p19 antigen near the range of cells that were not treated with any inhibitor. It is unclear why inhibitors at 20 μM are also in the same range as cells not treated and cells treated with the control. However, some of the inhibition seen may be due to cytotoxicity.

5.11 Conclusions

The MA/CA statine inhibitor has thus far shown the highest competitive inhibition based on fluorescence results with crude and purified compounds. The percent inhibition was found to be 88% compared to a previously reported CS-I-25 pepsin inhibitor, which had an 86% percent inhibition. The K_i for the MA/CA statine inhibitor was found to be 29 $\pm$ 4 nM, which is also close to the K_i = 30 $\pm$ 3 nM for CS-I-25. It was hoped that the statine, AHPPA, and HEA inhibitors would give better inhibition. One of the difficulties with HTLV-I protease inhibitor assays is that the protease is not well behaved in solution and the activity is low compared to other retroviral proteases. This may be a benefit to people who are infected with HTLV-I or are at risk for HTLV-I infection. If the protease were more active, the prevalence of HTLV-I infection and transmission might be much greater than the 15-20 million people infected that current statistics report.

REFERENCES

Akaji, K., Teruya, K. & Aimoto, S. 2003. Solid-phase synthesis of HTLV-1 protease inhibitors containing hydroxyethylamine dipeptide isostere. *J Org Chem*, **68,** 4755-63.

Daenke, S., Schramm, H. J. & Bangham, C. R. 1994. Analysis of substrate cleavage by recombinant protease of human T cell leukaemia virus type 1 reveals preferences and specificity of binding. *J Gen Virol*, **75,** 2233-9.

Ding, Y. S., Rich, D. H. & Ikeda, R. A. 1998. Substrates and inhibitors of human T-cell leukemia virus type I protease. *Biochemistry*, **37,** 17514-8.

Ha, J. J., Gaul, D. A., Mariani, V. L, Ding, Y. S., Ikeda, R. A., & Shuker, S. B. 2002. HTLV-I protease cleavage of p19/24 substrates is not dependent on NaCl concentration. *Bioorganic Chemistry* **30**(2), 138-144.

Hamada, T., Nishida, A. & Yonemitsu, O. 1986. Selective removal of electron-accepting r-toulene- and naphthalenesulfonyl protecting groups for amino function via photoinduced donor-acceptor ion pairs with electron-donating aromatics. *J. Am. Chem. Soc.*, **108,** 140-145.

Heneine, W., Yamamoto, S., Switzer, W. M., Spira, T. J. & Folks, T. M. 1995. Detection of reverse transcriptase by a highly sensitive assay in sera from persons infected with human immunodeficiency virus type 1. *J. Infectious Diseases*, **171,** 1210-1216.

Kitamura, K., Rudolph, D. L., Goldsmith, C., Folks, T. M. & Lal, R. B. 1993. Isolation, characterization, and transmission of human T-lymphotropic virus types I and II in culture. *curr. Microbiol.*, **27**.

Kuzmic, P., Sun, C. Q., Zhao, A. C. & Rich, D. H. 1991. Long range electrostatic effects in pepsin catalysis. *Tetrahedron*, **47,** 2519-2534.

Lal, R. B. & Heniene, W. 1996. Testing of human T-lymphotropic virus types I and II: serologic, virologic, and molecular detection. In: *Human T-Cell Lymphotropic Virus Type 1* (Ed. by Hafler, P. H. a. D.), pp. 167-195. West Sussex, England: John Wiley & Sons, Ltd.

Louis, J. M., Oroszlan, S. & Tözsér, J. 1999. Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type I proteinase. *J. Biol. Chem.*, **274,** 6660-6666.

Rich, D. H., Prasad, J. V., Sun, C. Q., Green, J., Mueller, R., Houseman, K., MacKenzie, D. & Malkovsky, M. 1992. New hydroxyethylamine HIV protease inhibitors that suppress viral replication. *J Med Chem*, **35,** 3803-12.

Segel, I. H. 1975. *Enzyme kinetics: Behavior and analysis of rapid equilibrium and steady-state enzyme systems*. New York: John Wiley and Sons.

Shuker, S. B., Mariani, V. M., Herger, B. E. & Dennison, K. J. 2003. Understanding HTLV-I Protease. *Chem. Biol.*, **5,** 373-380.

Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B. H. & Wlodawer, A. 1990. X-ray crystollographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. *Biochemistry*, **87,** 8805-8809.

Teruya, K., Kawakami, T., Akaji, K. & Aimoto, S. 2002. Total synthesis of L40I, C90A, C109A-human T-cell leukemia virus type 1 protease. *Tetrahedron Lett.*, **43,** 1487-1490.

Wellings, D. A. & Atherton, E. 1997. Standard Fmoc Protocols. *Meth. in Enzym.*, **289,** 44-66.

Zhao, T. M., Robinson, M. A., Bowers, F. S. & Kindt, T. J. 1995. Characterization of an infectious molecular clone of human T-cell leukemia virus type I. *J. Virol.*, **69,** 2024-2030.

CHAPTER 6

ACTIVITY AND STRUCTURE OF HTLV-I PROTEASE

6.1 Sequence Alignment

6.1.1 Mutation of HTLV-I Protease to Recognize an HIV-1 Substrate

Accurate alignment of retroviral proteases is difficult due to low sequence homology, especially since the structure of HTLV-I protease is unknown. In addition to our models, in previous work, Tözsér and others developed a theoretical model in order to determine which residues were significant for activity (Tözsér et al., 2000). A particular method for testing alignment accuracy is to mutate residues of the unknown protein aligned with that of a known structure and test for specificity. Wu et al. used this strategy to mutate RSV protease residues to match those of HIV-1 protease (Wu et al., 1998). Employing that same strategy, Bryan Herger mutated HTLV-I protease to match putative active site residues in HIV-1. The alignments of RSV and HIV-1 protease along with three HTLV-I alignments are shown in Figure 6.1. The three HTLV-I aligments are: Tözsér's (T), our previously published alignment (S1), and our revised alignment (S2). The RSV wt residues highlighted in yellow indicate those that were mutated by Wu et al.

```
HTLV-T  PVIPLDPARRPVIKAQVDTQT-----S-HPKTIEALLDTGADMTVLPIALFSS-NT-P-L-KN-T--SVLGA
HTLV-S1 PVIPLDPARRPVIKAQVDTQT---S-HP--KTIEALLDTGADMTVLPIALFSSN-----TP-LK--N-TSVL
HTLV-S2 PVIPLDPARRPVIKAQVDTQT---S-HP--KTIEALLDTGADMTVLPIALFSSN-----TP-LKN-TSVLGA
RSVwt   LAMTMEHKDRPLVRVILTNTGSHPVKQ-RSVYITALLDSGADITIISEEDWPT-DW-P-V-MEAANPQIHGI
HIV-1   PQITLW--KRPLVTIKIG-----------GQLKEALLDTGADDTVIEE--MS---LPGRW-KPK---MIGGI

HTLV-T  GGQTQDHFK-LTSLPVLIRLPFRTTPIVLTSCLVDTKNNWA--IIGRDALQQCQGVLYLPEAKGPPVIL
HTLV-S1 GAGGQTQDHFK-LTSLPVLIRLPFRTTPIVLTSCLVDTKNNWAIIGRDALQQCQGVLYLPEAKGPPVIL
HTLV-S2 GGQTQDHFK-LTSLPVLIR--LPFRTTPIVLTSCLVDTKNNWAIIGRDALQQCQGVLYLPEAKGPPVIL
RSVwt   GGGIPMRKSRD-MIELGVINRDGSLERPLLLFPAVAMVRGS--ILGRDCLQGLGLRLTNL---------
HIV-1   GGFIKV-RQY-DQIIIEIC-----GH-KAIGTVLVGPTPVN--IIGRNLLTQIGCTLNF----------
```

Figure 6.1 Multiple alignment of RSV protease (RSVwt), HTLV-I protease (Shuker alignment: S), HTLV-I protease (Tozser alignment: T), HIV-1 protease. Residues in yellow in RSV protease were mutated by Wu et al. to the corresponding residues in HIV-1 protease.

The mutations based on these alignments are shown in Figure 6.2(a,b,c). The mutations based on the Tözsér alignment were: M37D, D65V, V92L, N96T, N97P, W98V, and A99N. The mutations based on the S1 alignment were: M37D, T63V, C90L, K95P, and N96V. The mutations based on the S2 alignment were: M37D, D65V, C90L, K95P, and N96V.

(a)

```
                                        D
PVIPLDPARRPVIKAQVDTQTSHPKTIEALLDTGADMTVLPIALFSSNTPLKNTSVLGAGGQTQ
DHFKLTSLPVLIRLPFRTTPIVLTSCLVDTKNNWAIIGRDALQQCQGVLYLPEAKGPPVIL
V                          L   TPVN
```

(b)

```
                                        D                           V
PVIPLDPARRPVIKAQVDTQTSHPKTIEALLDTGADMTVLPIALFSSNTPLKNTSVLGAGGQTQ
DHFKLTSLPVLIRLPFRTTPIVLTSCLVDTKNNWAIIGRDALQQCQGVLYLPEAKGPPVIL
                           L    PV
```

(c)

```
                                        D
PVIPLDPARRPVIKAQVDTQTSHPKTIEALLDTGADMTVLPIALFSSNTPLKNTSVLGAGGQTQ
DHFKLTSLPVLIRLPFRTTPIVLTSCLVDTKNNWAIIGRDALQQCQGVLYLPEAKGPPVIL
V                          L    PV
```

Figure 6.2 HTLV-I protease mutations made based on different alignments. Target residues are bold with mutations above or below the parent sequence. (a) mutations based on Tözsér (T) alignment; (b) Mutations based on Shuker alignment S1; (c) mutations based on Shuker alignment S2.

6.1.2 Kinetic Assays of HTLV-I Protease Mutants

For the kinetic assays of HTLV-I mutant constructs, Bryan Herger used an HIV-1 substrate (H-5910) to measure cleavage based changes in absorbance (Toth & Marshall, 1990). His results are summarized in Table 6.1.

Table 6.1 Summary of kinetic data for the various protease mutants.

Mutant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (mM^{-1} s^{-1})
HIV-1 protease (Toth & Marshall, 1990)	0.29 ± 0.03	37 ± 8	7.8 ± 0.3
HTLV-I protease: combined mutant (S2)	$3.6x10^{-3} \pm 2.9x10^{-4}$	64 ± 10	0.056 ± 0.0045
HTLV-I protease: Shuker mutant (S1)	$1.6x10^{-3} \pm 1.3x10^{-4}$	52 ± 8	0.031 ± 0.0025
HTLV-I protease: Tözsér mutant (T)	$4.5x10^{-4} \pm 4.4x10^{-5}$	39 ± 8	0.012 ± 0.0011
HTLV-I protease: wild type	Not detected	Not detected	Not detected

The mutant based on Tözsér's alignment was shown to have a catalytic efficiency, k_{cat}/K_M, of 0.012 mM^{-1} s^{-1}, while the mutant based on Shuker's alignment gave a k_{cat}/K_M of 0.031 mM^{-1} s^{-1}. The S2 mutant, which had two mutations from Tözsér's alignment and two from S1 had a k_{cat}/K_M of 0.056 mM^{-1} s^{-1}, indicating that Asp^{65} plays a more significant role in binding substrate. These data indicate that the revised S2 alignment better correlates active site residues in HIV-1 and HTLV-I proteases. Therefore, the theoretical model based on this alignment is more likely to reflect the actual structure of HTLV-I protease based on similarity to HIV-1 protease. The second theoretical model was deposited in the protein data bank based on this revised alignment. This model is shown in Figure 6.3 (Herger et al., 2004). Figure 6.4 shows the two theoretical models, PDB 1O0J and PDB 1TP1, superimposed. The superimposed

illustration shows that there are marked differences in the two structures, especially in the exterior flap regions.

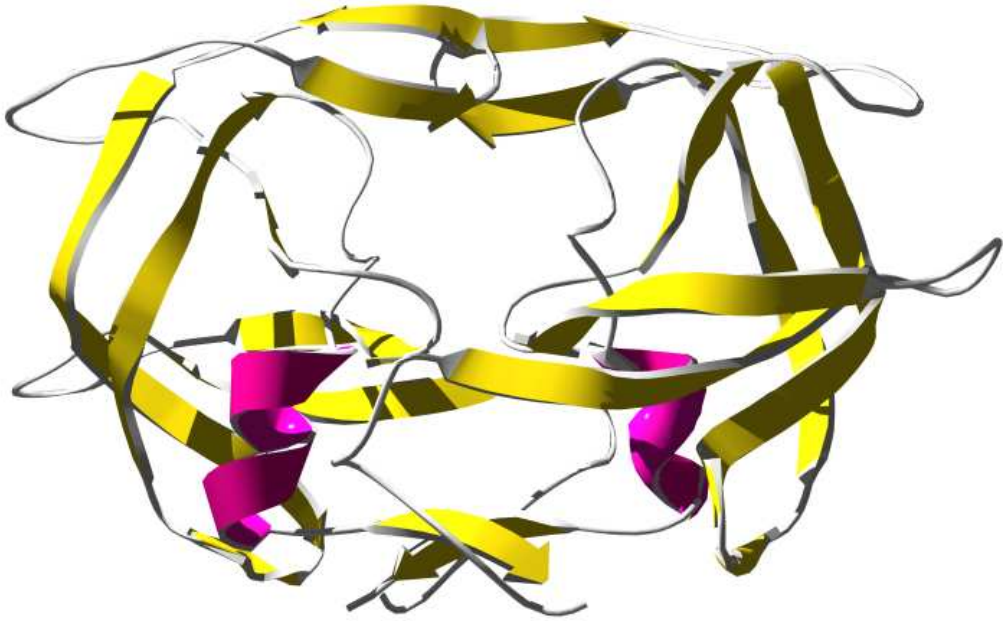


Figure 6.3 Theoretical model PDB 1TP1 based on Shuker (S2) revised alignment.

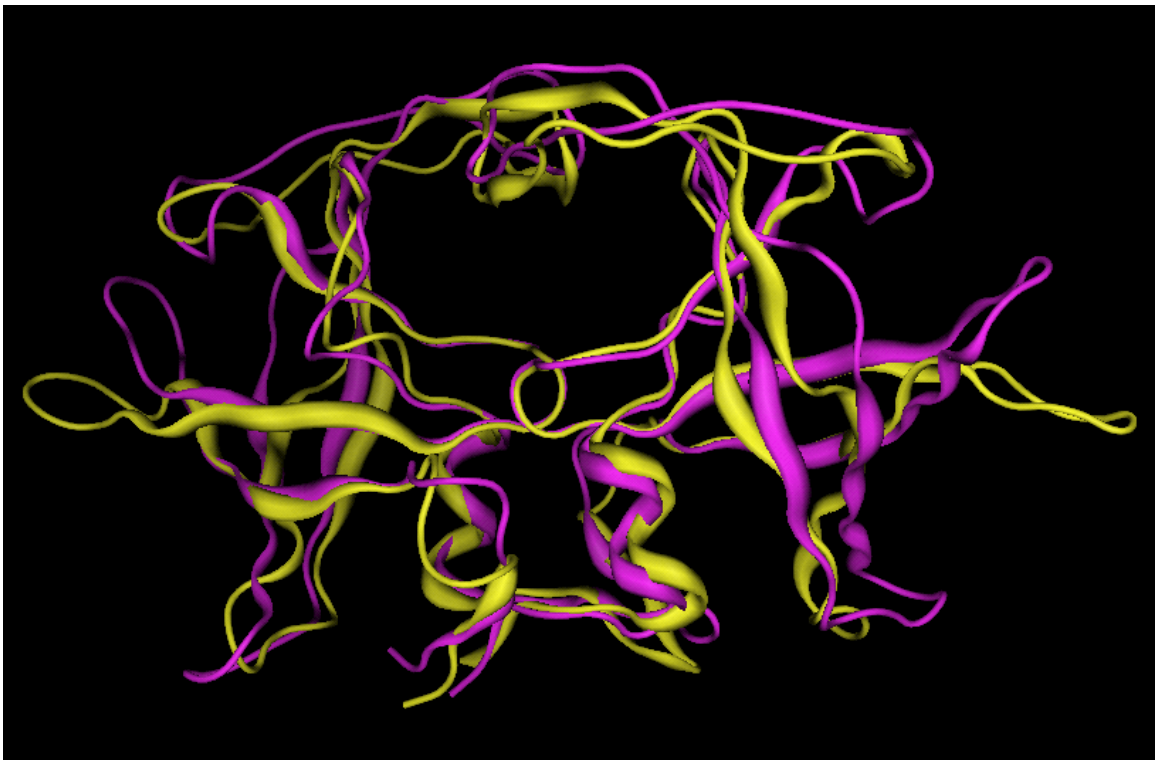


Figure 6.4 Superimposed thereotical models. PDB 1O0J (gold) and PDB 1TP1 (magenta).

6.2 Location of Tryptophan Residues in HTLV-I Protease Theoretical Models

Unlike other retroviral proteases, HTLV-I protease contains a tryptophan residue at position 98 in each dimer. This position is in proximity to residues Ile^{100}-Ile^{101}-Gly^{102}, which are in the active site. However, without a known structure it is difficult to determine where those two tryptophan residues lie based solely on sequence alignment. In the first

theoretical model deposited in the protein data bank, PDB 1O0J, the two tryptophans are located on the exterior of the protein. This is illustrated in Figure 6.5 where the Trp^{98} is shown as purple sticks in a top down view of the protease. This first model was based on our first (S1) homology alignment (Herger, 2004; Shuker et al., 2003).

Figure 6.5 Top down view of PDB 1O0J with tryptophans shown in purple.

In the second theoretical model, PDB 1TP1, based on our revised (S2) alignment (Herger et al., 2004) the tryptophans are also located on the exterior of the protease. Figure 6.6 shows a top view of PDB 1TP1 with the tryptophans shown in purple. There are some rotational differences between the two tryptophans in each structure. Both are consistent with displaying the tryptophans as solvent exposed. However, in an earlier unpublished model based on the S1 alignment, the tryptophans are in the interior of the

protein. This model was also prepared using MODELLER. Figure 6.7 shows this model with the buried tryptophans shown in purple.

Figure 6.6 Top down view of PDB 1TP1 with tryptophans shown in purple.

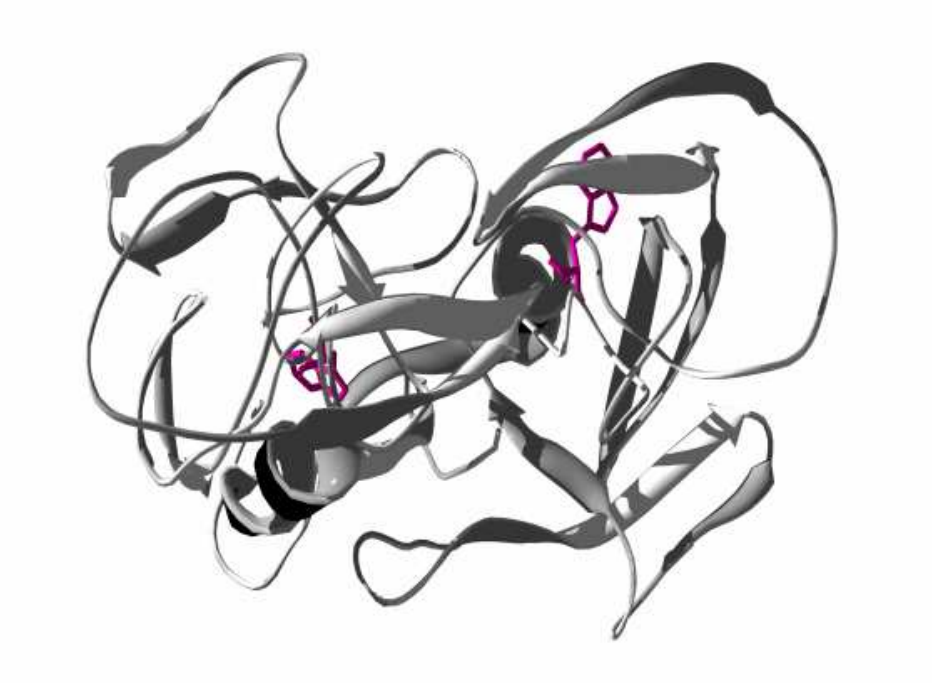


Figure 6.7 Unpublished theoretical model of HTLV-I protease (S1 alignment) with tryptophans shown in purple.

The question remained as to which of these structures most acurately depicts the location of the tryptophan residues. It is expected that the large hydrophobic tryptophans would reside in the interior of the protein.

6.3 Alanine Scan of HTLV-I Protease

Recognizing which residues in an active site domain contribute to catalytic efficiency is important for understanding the structure and function of an enzyme. An alanine scan can be a critical diagnostic tool for identifying these residues (Cunningham & Wells, 1989). This technique involves mutating a putative active site residue to an alanine and then testing catalytic activity. Alanine is chosen because it is the smallest chiral amino acid. Bryan Herger prepared several constructs containing alanine mutations of sites that were speculated to be important for activity. These mutants were: Arg^{10}, Leu^{30}, Thr^{94}, Lys^{95}, Asn^{96}, Trp^{98}, and Ile^{100}. Kinetic assays were performed with the MA/CA subrate analogue, APQVL(Nph)VMHPL, where Nph is 4-nitrophenylalanine. Herger's results are shown in Table 6.2.

Table 6.2 Kinetic Results of Alanine Mutants.

Mutant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})	Whole number ratio
L40I	2.74 x 10^{-3}	43.2	63.5	1
R10A	7.18 x 10^{-3}	61.6	117	2
L30A	5.72 x 10^{-3}	71.6	79.9	1
T94A	3.01 x 10^{-3}	111	27.2	-2
K95A	9.36 x 10^{-3}	73.6	127	2
N96A	6.03 x 10^{-3}	70.3	85.7	1
W98A	2.96 x 10^{-3}	8.33	355	6
I100A	4.43 x 10^{-3}	43.3	102	2

The k_{cat}/K_M is the apparent second-order rate constant for the enzyme to catalyze transformation of the substrate into product. Therefore, it can be used as a measure of catalytic efficiency for the enzyme (Garrett & Grisham, 1995). The threonine mutation was the only one to reduce catalyitic activity with a k_{cat}/K_M of 27.2 $M^{-1}s^{-1}$. Thr^{94} may have a stabilizing function through hydrogen bonding to the P1 leucine of the substrate. The Ile^{100}, Arg^{10}, and Lys^{95} all increased the activity with k_{cat}/K_M of 102 $M^{-1}s^{-1}$, 117 $M^{-1}s^{-1}$, 127 $M^{-1}s^{-1}$, respectively. The Trp^{98} mutation had a dramatic effect in increased protease activity with a k_{cat}/K_M of 355 $M^{-1}s^{-1}$. The K_M value of 8.33 μM was the major contributor to this marked increase in activity. The K_M is inversely proportional to the enzyme affinity for its substrate (Garrett & Grisham, 1995). Therefore, the loss of the tryptophans gives

the protease a higher affinity for the MA/CA substrate. This is significant because it suggests that the tryptophans must be located near the active site of the protein. Unlike other retroviral proteases, HTLV-I protease contains tryptophan residues at positions in proximity to residues Ile100-Ile101-Gly102, which are in the active site. The question still remained as to whether or not the tryptophans flank the active site and are solvent exposed or are actually buried in the flaps of the active domain.

6.4 Fluorescent Quenching of HTLV-I Protease Tryptophan[98]

A common method to determine the location of tryptophans is to use fluorescence quenching of the tryptophan residues. In proteins, tryptophan is typically the most dominant fluorophore present with an absorbance maximum near 280 nm. The emission maximum of tryptophan is very sensitive to the polarity of the surrounding environment. A tryptophan that is buried deep inside a protein has an emission maximum near 320 nm where as a solvent-exposed tryptophan has an emission maximum near 350 nm.

6.4.1 Results from Quenching Experiments

In acrylamide quenching experiments, a linear quenching of the tryptophan fluorescence peak is expected due to the ability of acrylamide to penetrate the interior of the protein in its native state. For KI quenching, there should be no reduction in the fluorescence peak for buried tryptophan residues in the native protein. However, once denatured, a reduction in the fluorescence tryptophan peak is expected because all tryptophan residues should be exposed.

The effects of quenching of native HTLV-I tryptophans with acrylamide are shown in Figure 6.8. The most intense curve resulted in the absence of quencher. Subsequent dilutions with acrylamide led to considerable quenching at 350 nm. Because acrylamide is hydrophobic, it is able to penetrate the folded protein and interact with the tryptophan residues allowing for adequate quenching of the fluorophores. Dilutions of KI into the native protease solution were carried out (Figure 6.9). For this case, the least intense curve is the one with no KI added. Rather than quenching, there is an increasing fluorescence with increasing KI due to the fluorescence of iodine. The increase in signal upon KI additions can be attributed to the free I_3^- in solution, which has an absorbance in the wavelength region of tryptophan residues. When the protease was denatured however, we saw considerable quenching as is shown in Figure 6.10. Again, the most intense curve resulted is without any KI and subsequent quenching can be seen as higher concentrations of KI are added to the solution. From quenching the experiments with KI, the fact that KI does not quench the tryptophans in the native protease, but does

quench the denatured protease, indicates that the two tryptophans are buried within the protease interior.

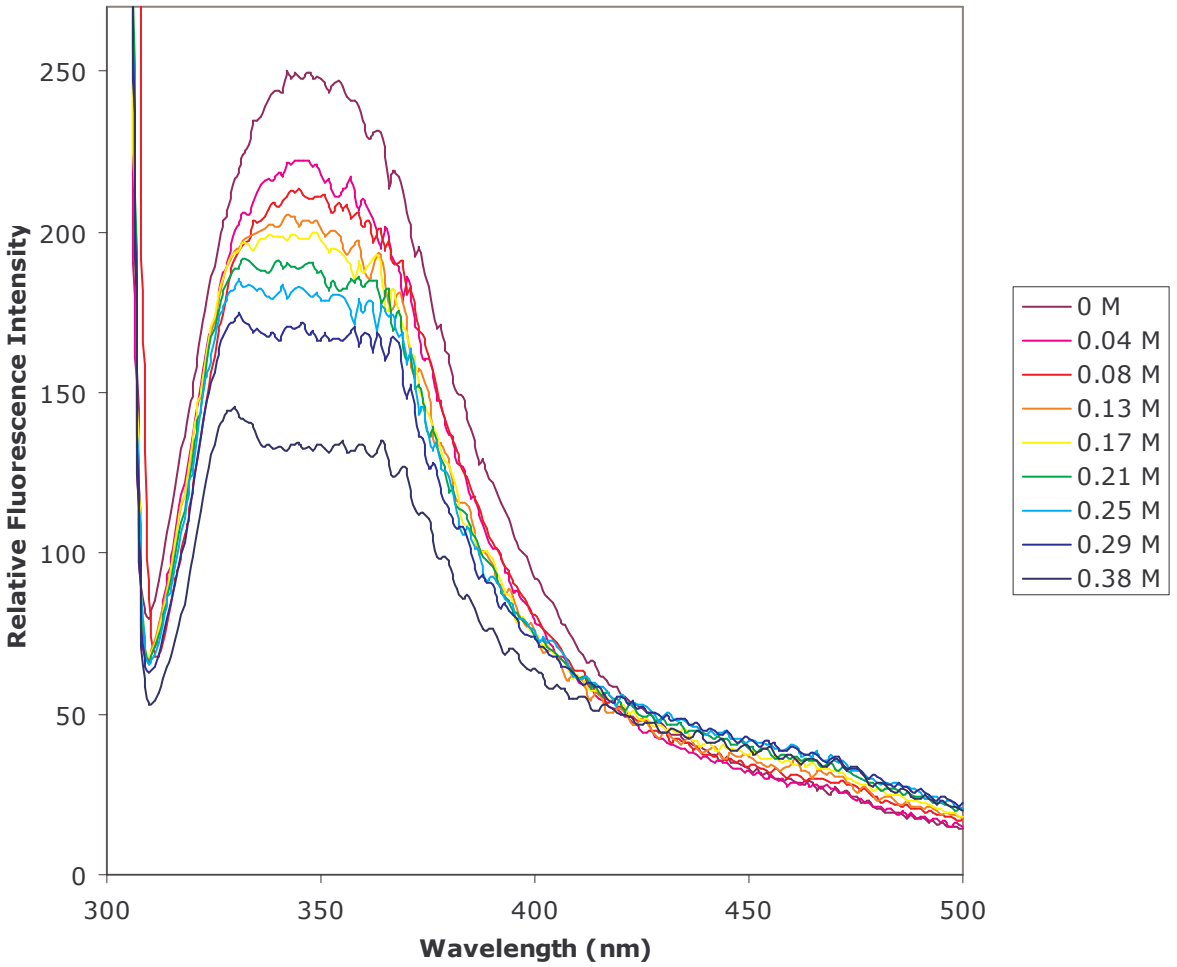


Figure 6.8 Fluorescence quenching of native HTLV-I protease tryptophan residues with acrylamide.

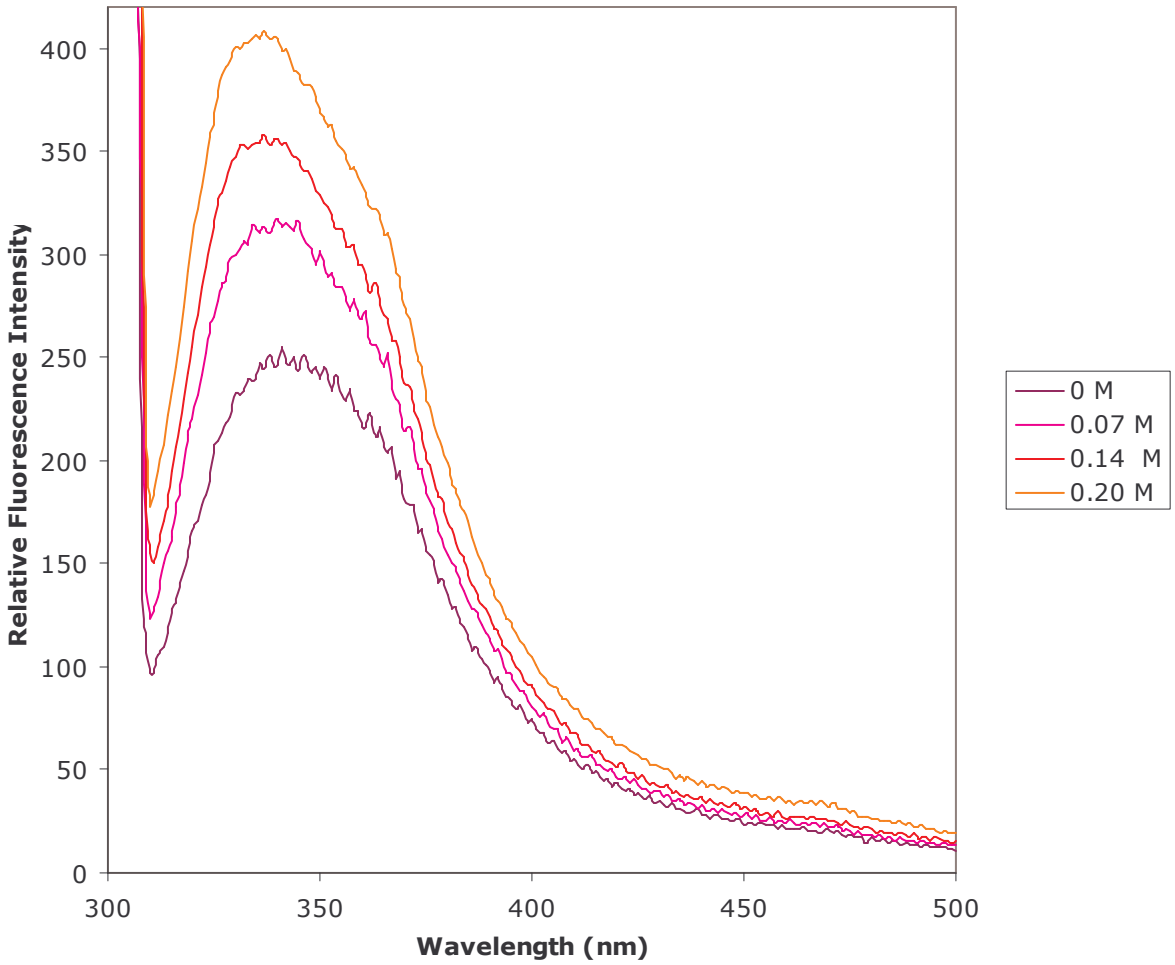


Figure 6.9 Fluorescence quenching of native HTLV-I tryptophan residues with KI.

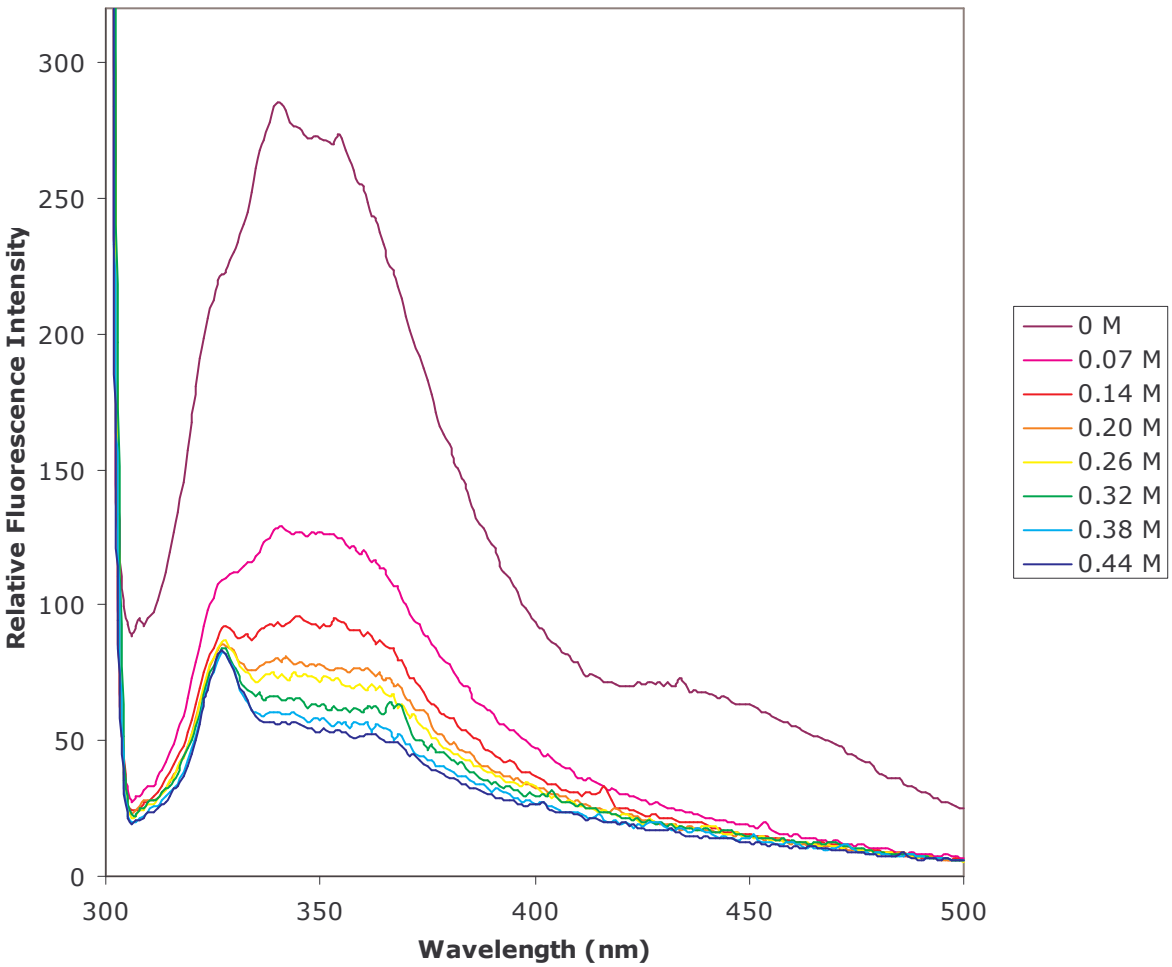


Figure 6.10 Fluorescence quenching of denatured HTLV-I protease tryptophan residues with KI.

A quantitative analysis of collisional fluorescence quenching can be carried out according to the modified Stern-Volmer equation (Yi-Brunozzi et al., 2001; Weber & Senior, 2000; Lehrer, 1971; Eftink & Ghiron, 1977; Eftink & Ghiron, 1976),

$$\frac{F_0}{F} = K_{SV}[Q] \qquad \text{(Eq. 1)}$$

and F are the fluorescence intensities in the absence and presence of quencher, respectively, [Q] is the concentration of quencher, V is the static quenching constant, and K_{sv} is the dynamic Stern-Volmer constant. The fluorescence intensity can be written in terms of the sum of fluorophores exposed to solvent,

$$F_0 = F_{0,a} + F_{0,b} \qquad \text{(Eq. 2)}$$

where $F_{o,a}$ and $F_{o,b}$ are the individual initial intensities related to each species. The fraction of the initial fluorophores accessible to quencher can be determined using a modified Stern-Volmer equation (Eftink & Ghiron, 1977; Eftink & Ghiron, 1976; Yi-Brunozzi et al., 2001),

$$\frac{F_0}{F_0 - F} = \frac{1}{[Q] f_a K_Q} + \frac{1}{f_a} \qquad \text{(Eq. 3)}$$

where f_a is the fraction of accessible fluorophores. For nonlinearity, our data were fitted using PSI-Plot (Poly Software International) to distinguish between the sphere-of-action

static quenching and multiple fluorophore models. The static sphere-of-action quenching model was derived from the equation,

$$\frac{F_0}{F} = 1 + K_{SV}[Q]e^{V[Q]} \qquad \text{(Eq. 4)}$$

where V is the static quenching component (Weber & Senior, 2000; Yi-Brunozzi et al., 2001). For multiple fluorophores (i.e. two tryptophans, one per monomer) where there are two equivalent but noninteracting species present, the following equation was applied,

$$\frac{F_0}{F} = \frac{1 + K_{SV}^a[Q]}{\left(1 + K_{SV}^a[Q]\right)\left(1 - f_a\right) + f_a} \qquad \text{(Eq. 5)}$$

where, $$f_a = \frac{F_0^a}{F_0} \qquad \text{(Eq. 6)}$$

relates the fraction of initial fluoresence emitted by one species present (see Eq. 3) with its Stern-Volmer constant: K_{SV}^a.

Even though it appears that the tryptophans are internal by qualitative analysis, analysis of the fluorescence at 350 nm using the linear Stern-Volmer by plotting equation F_o/F vs. [KI] or [acrylamide] (Eq. 1) did not give linear results. As shown in Figure 6.11, acrylamide quenching resulted in a concave downward slope and KI quenching gave a concave upward slope. Nonlinear Stern-Volmer plots indicate that there is heterogeneity among fluorophores.

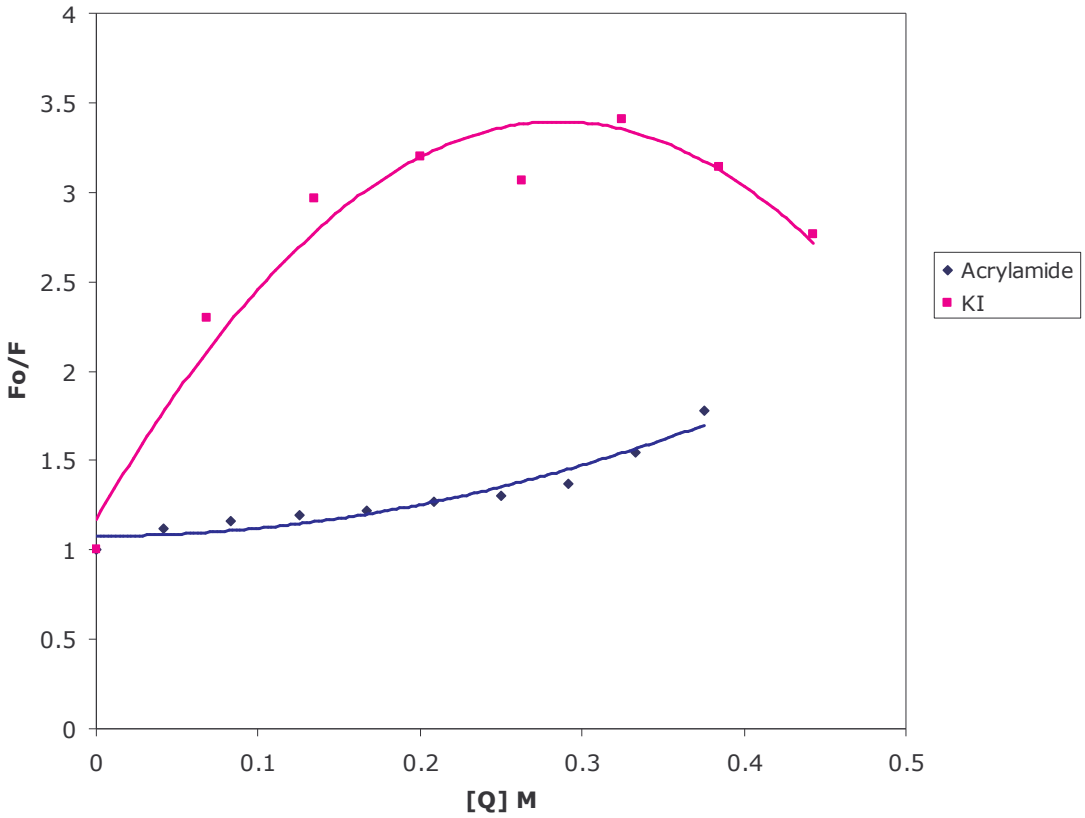


Figure 6.11 Relative corrected intensity ratio (F_o/F) vs. quencher concentration ([Q]) in M). The upper curve (pink ■) represents the quenching denatured protease with KI and the lower curve (blue ♦) represents acrylamide quenching of native protease.

Deviation from linearity will occur due to multiple fluorescing centers. They results here suggest that the two tryptophans in the dimer have varying degrees of accessibility to the quencher. In HTLV-I protease, the tryptophan residue of one monomer may be more buried than the tryptophan on the other.

Employing the modified Stern-Volmer equation, $F_o/(F_o - F)$ vs. $1/[Q]$ (Eq. 3) was plotted (Figure 6.12) where the reciprocal of the y-intercept is equal to f_a, which pertains to the fraction of accessible fluorophores. This plot yielded fractions of 80% accessible fluorophores for KI quenching and 92% for acrylamide. The fraction of accessible

fluorophores for KI quenching of the denatured enzyme should be equal to 1 because all tryptophan residues would be accessible to KI in the denatured state. It is not known why these data show only 80% accessibility.

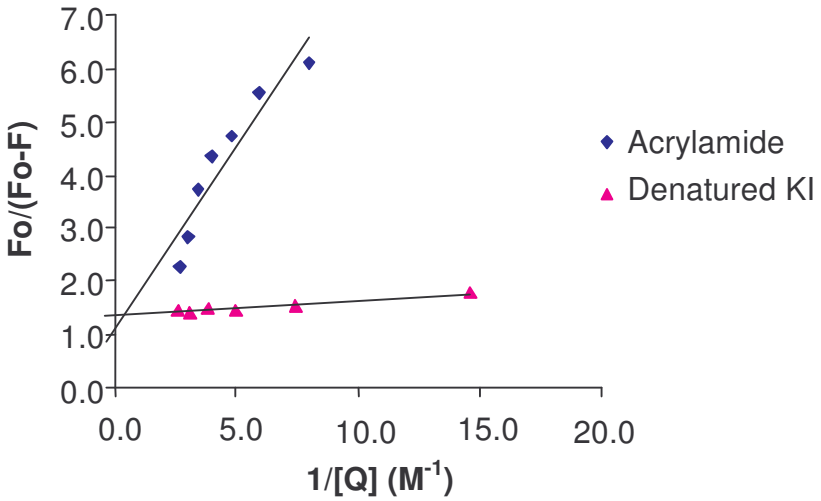


Figure 6.12 Modified Stern-Volmer plots for acrylamide quenching of folded HTLV-I protease (blue ♦) and KI quenching of denatured protease (pink ▲).

The 92% accessibility of the tryptophans to acrylamide quenching suggests that the fluorophores do not behave homogeneously and most likely have slight differences in orientation. In fact, HTLV-I protease is probably not a truly C2 symmetric homodimer. Other retroviral proteases such as HIV-1 and RSV are known to fit pseudo-C2 point groups when binding substrate and possess an odd number of waters in the active site

(Tözsér et al., 2000; Prabu-Jeyabalan et al., 2000). It may be that retroviral proteases in the absence of substrate are asymmetric dimers *in vivo*. In this case, based on sequence alignment with other aspartyl proteases, HTLV-I would also confer a pseudo-C2 symmetry. If so, it is highly likely that the two tryptophans have significant translational differences resulting in equivalent but non-interacting fluorescence by these residues. Using PSI-Plot (Poly Software International) the data were fitted to the sphere-of-action static quenching model (Eq. 4) for KI and to the multiple fluorophore model (Eq. 5) for acrylamide. The linear results agree with these phenomena as illustrated in Figure 6.13.

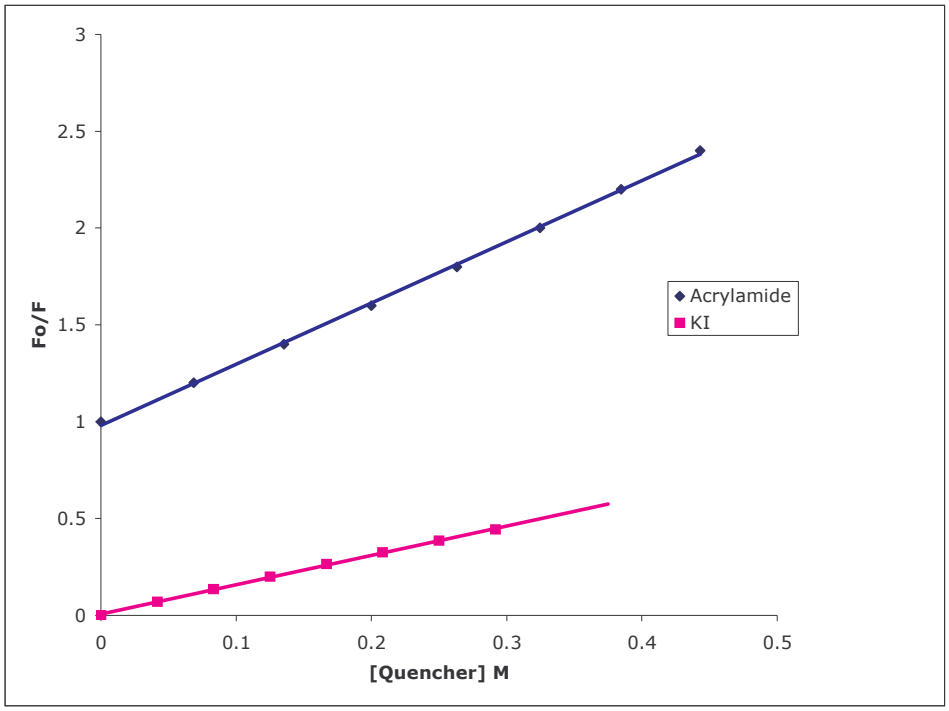


Figure 6.13 Plot employing fitted multiple fluorophore model for acrylamide quenching of folded HTLV-I protease and static sphere-of-action model for KI quenching of denatured protease.

The data from KI and acrylamide were fitted to both Eqs. 4 and 5. The best-fits were found to apply to static quenching for KI and multiple fluorophore model for acrylamide quenching.

The K_{SV} can be found for acrylamide quenching from the plot in Figure 6.13, where the slope is equal to $1/f_a K_{SV}$. The K_{SV} for acrylamide quenching was found to be 1.96 M^{-1}. This is in agreement with the range of several other enzymes shown in Table 6.3 (Eftink and Ghiron, 1977).

Table 6.3 Acrylamide Quenching K_{SV} for Several Enzymes.

Enzyme	**K_{sv} (M^{-1})**
Pepsin	9.5
Nuclease	5.2
β-Trypsin	2.4
R Nase T_1	1.1

6.4.2 Conclusions from Quenching Studies

From the fluorescence results, the acrylamide and KI quenching showed deviations from linearity, which imply that there are multiple fluorescing centers. The acrylamide quenching on native protease was fitted to the multiple fluorophore model

and yielded 92% accessibility to the quencher. This suggests slight rotational differences between the two tryptophans (one per dimer).

From the KI quenching results on both native and denatured protease, it can be concluded that the theoretical models deposited in the protein data bank are inaccurate representations due to the location of the Trp^{98} residues. The lack of quenching by KI in the native protease and positive quenching in denatured protease indicate that the tryptophan residues are indeed in the interior of the protein. Theoretical models are important tools for better understanding of structure when three-dimensional crystal structures are not available. However, theoretical models must be constantly updated as more information becomes available regarding structure.

Of major significance in this study is that quenching by KI provides a quick analytical tool for determining whether or not HTLV-I protease is folded. Before NMR samples are prepared, KI quenching experiments can be run on protease samples to validate whether the protease is suitable for structure experiments.

6.5 Structure Determination by NMR

The three dimensional structure of HTLV-I protease is critical for accurate representation of enzymology. However, to date no X-ray or NMR structure has been reported. It has been a primary goal of the Shuker group to obtain a three-dimensional structure of HTLV-I protease. The work of structure determination has been in collaboration with Dr. Alex Wlodawer at the National Cancer Institute (Frederick, MD) for

the X-ray structure, Dr.'s Vicky Bevilacqua and Jeffrey Rice at Edgewood Chemical Biological Center (Aberdeen Proving Ground, MD) and Mike Goeger of the New York Structural Biology Center (NYSBC) (New York, NY) for the NMR structural work.

6.6 NMR Spectra of HTLV-I Protease

Below in Figures 6.14 and 6.15 are 800 MHz NMR spectra obtained at the NYSBC. These data are similar to those seen in previous experiments carried out by David Gaul, Bryan Herger, Mike Kulis, and myself in the Shuker laboratory on ^{15}N-labeled protease.

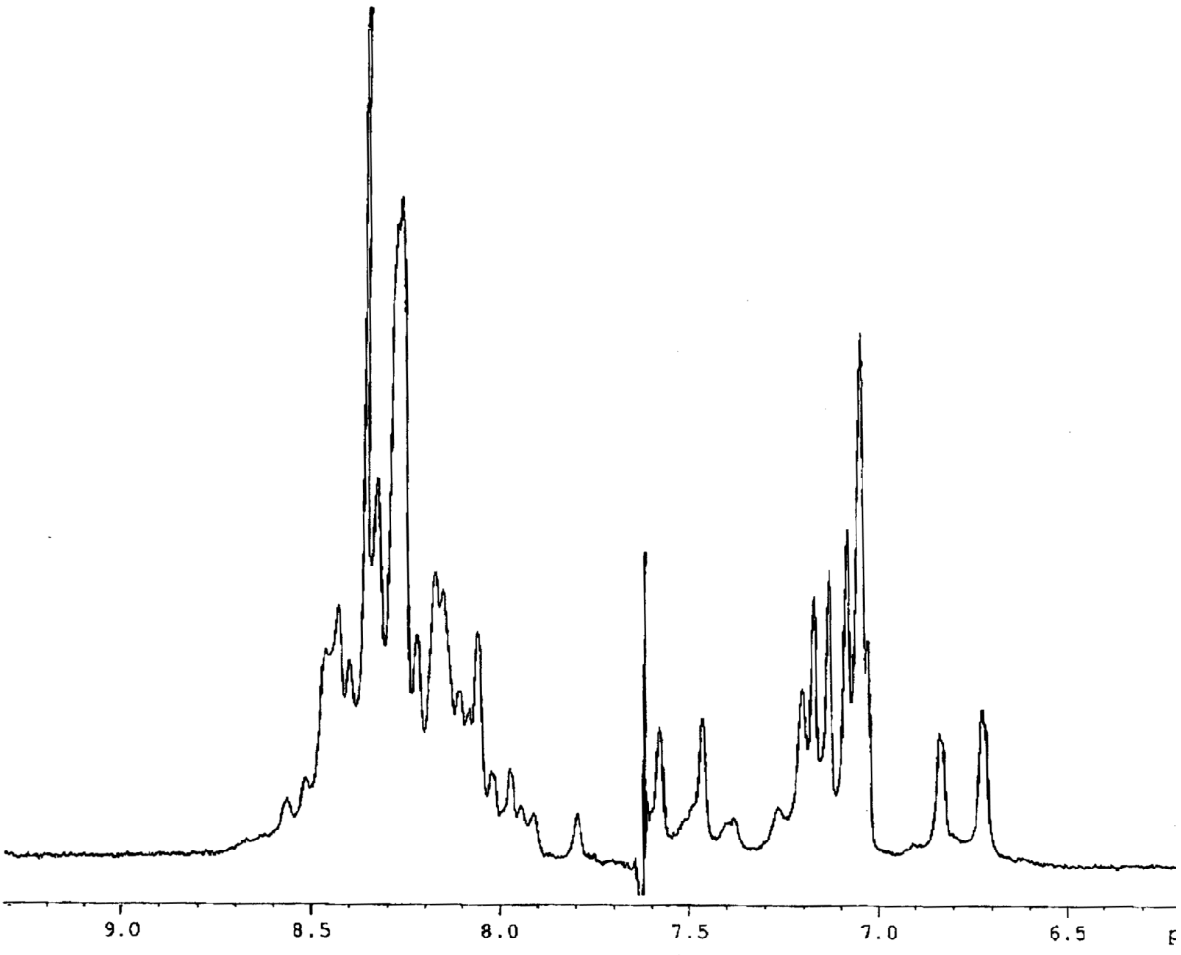


Figure 6.14 HTLV-I protease 800 MHz 1D ^{1}H NMR amide (100 μM [protease], 50 mM sodium acetate, 1 mM DTT, and pH 5.3).

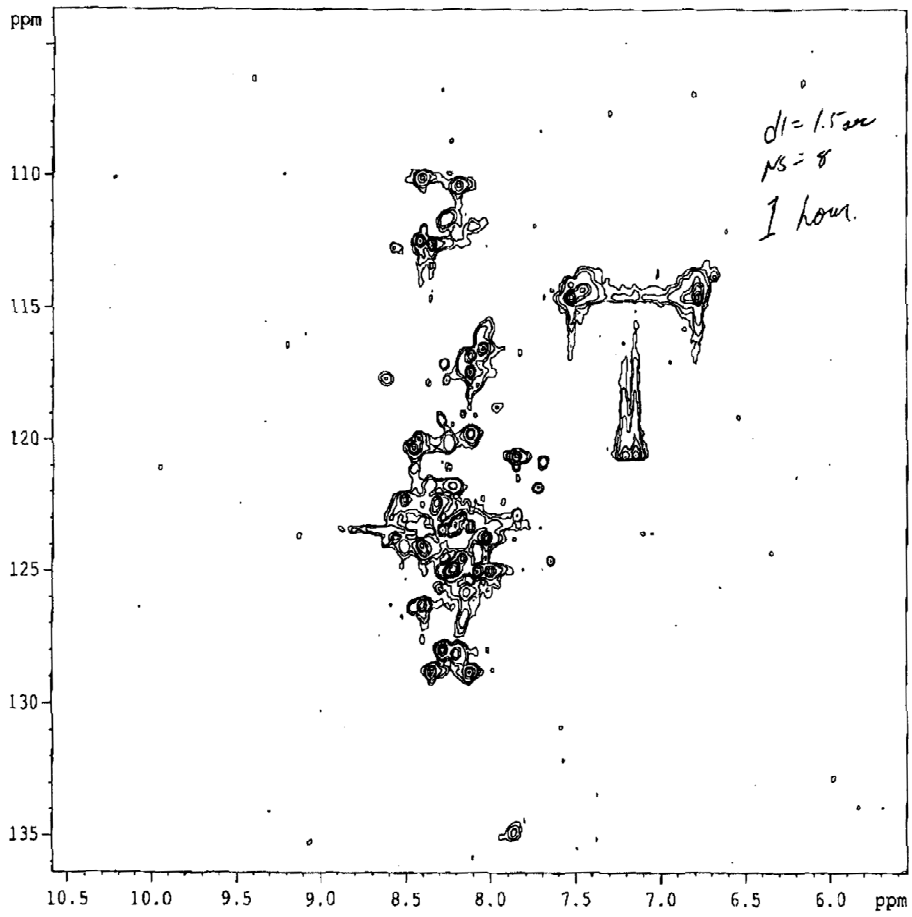


Figure 6.15 HTLV-I protease HSQC after 1 hour of processing (100 μM [protease], 50 mM sodium acetate, 1 mM DTT, and pH 5.3).

A properly folded protein of this molecular weight should a number of well-resolved amide peaks above 8.5 parts per million (ppm) and below 6.5 ppm. These peaks evident in the 1D spectrum and the related chemical shifts would result in peaks spread throughout the HSQC spectrum. The lack of such resolution in the spectra shown here is primarily occupying a random coil

structure under these conditions (100 μM [protease], 50 mM sodium acetate, 1 mM DTT, and pH 5.3).

6.7 Conclusions

There are significant problems that must be addressed in order to obtain a correctly folded enzyme. First, HTLV-I protease is comprised of mostly hydrophobic residues. There are expected to be many phenylalanine residues exposed to the exterior of the protein. These may complicate the structure by facilitating aggregation. Second, the issue of properly folding the protein has not been resolved. The protease remains active in a partially folded state so that kinetic experiments demonstrate definable differences between various substrates and inhibitors. However, as revealed from repeated KI quenching experiments prepared following the same conditions, the protease is only folded 10-20% of the time. In addition, it appears from the MALDI results for the C-terminal cleavage data, that there is an additional autoprocessing site at Leu78. The autoprocessing of this site may result in the denaturing of the protease, which may account for the random coil NMR data.

The main step toward obtaining an NMR structure is to have adequate amounts of labeled protein. We were able to express enough ^{15}N-labeled protein so that an HSQC spectrum with reasonable signal-to-noise could be obtained in approximately one hour. Perhaps, an L40I-L78I mutant can be ^{15}N-labeled and

expressed. This may be the next step toward having HTLV-I protease stable enough to obtain folded data.

REFERENCES

Cunningham, B. C. & Wells, J. A. 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, **244,** 1081-5.

Eftink, M. R. & Ghiron, C. A. 1976. Exosure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching. *Biochemistry*, **15,** 672-680.

Eftink, M. R. & Ghiron, C. A. 1977. Exposure of tryptophanyl residues and protein dynamics. *Biochemistry*, **16,** 5546-5551.

Garrett, R. H. & Grisham, C. M. 1995. *Biochemistry*. Orlando: Saunders College Publishing.

Herger, B. E., Dennison, K. J. & Shuker, S. B. 2005. Identifying active site residues of HTLV-I protease: Mutation of HTLV-I protease to recognize an HIV-1 protease substrate. *Arch. Virol.* **submitted**.

Herger, B. E., Mariani, V. M., Dennison, K. J., Shuker, S. B. 2004. The Ten C-Terminal Residues of HTLV-I Protease Are Not Necessary for Enzymatic Activity. *Biochemical and Biophysical Research Communications*, **320,** 1306-1308.

Lehrer, S. S. 1971. Solute perturbation of protein fluorescence. The quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. *Biochemistry*, **10,** 3254-3263.

Prabu-Jeyabalan, M., Nalivaika, E. & Schiffer, C. A. 2000. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.*, **310,** 1207-1220.

Shuker, S. B., Mariani, V. M., Herger, B. E. & Dennison, K. J. 2003. Understanding HTLV-I Protease. *Chem. Biol.*, **5,** 373-380.

Toth, M. V. & Marshall, G. R. 1990. A simple, continous fluorometric assay for HIV protease. *Int. J. Pept. Protein. Res.*, **36,** 544-550.

Tözsér, J., Zahuczky, G., Bagossi, P., Louis, J. M., Copeland, T. D., Oroszlan, S., Harrison, R. W. & Weber, I. T. 2000. Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. *Eur. J. Biochem.*, **267,** 6287-6295.

Weber, J. & Senior, A. E. 2000. Features of F1-ATPase catalytic and noncatyltic sites revealed by fluorescence lifetimes and acrylamide quenching of specificity inserted tryptophan residues. *Biochemistry*, **39**.

Wu, J., Adomat, J. M., Ridky, T. W., Louis, J. M., Leis, J., Harrison, R. W. & Weber, I. T. 1998. Structural basis for specificity of retroviral proteases. *Biochemistry*, **37,** 4518-26.

Yi-Brunozzi, H. Y., Stephens, O. M. & Beal, P. A. 2001. Conformational changes that occur during an RNA-editing adenosine deamination reaction. *J. Biol. Chem.*, **276,** 37837-37833.

CHAPTER 7

HTLV-I NUCLEOCAPSID

7.1 Nucleocapsid Proteins

All retroviruses, except spumaviruses, have nucleocapsid proteins that contain one or two zinc fingers with a highly conserved $CX_2CX_4HX_4C$ (CCHC) motif where C = cysteine, H = histidine, and X = any amino acid (Amarasinghe *et al.*, 2000; Coffin *et al.*, 1997; Lee *et al.*, 1998; Williams *et al.*, 2002). Each CCHC binds a zinc atom as illustrated in Figure 7.1. This figure shows HIV nucleocapsid structure with the bound zinc atoms.

The nucleocapsid is cleaved from the Gag polyprecursor and has its primary function in packaging the viral genome (Bertola *et al.*, 2000; Coffin *et al.*, 1997; Stephen *et al.*, 2002). Within the Gag precursor, the nucleocapsid domain is involved in both recognition and packaging the RNA. Figure 7.2 shows a typical retroviral nucleocapsid CCHC interaction with RNA.

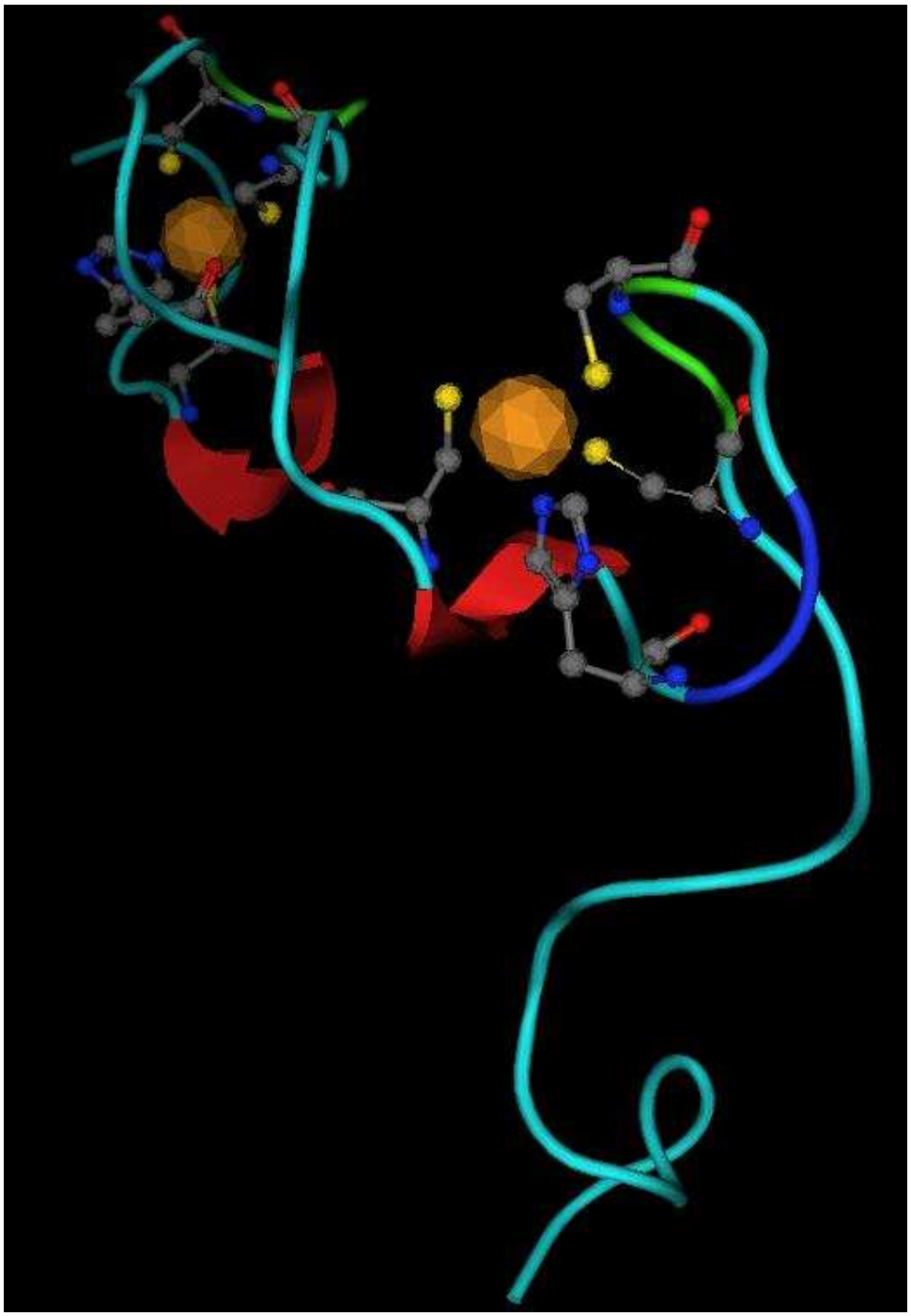


Figure 7.1 HIV nucleocapsid (PDB 1mfs) Lee, *et al.* (1998) *J. Mol. Biol*. 279, 633-649. Orange spheres = zinc atoms, yellow balls = cysteine residues.

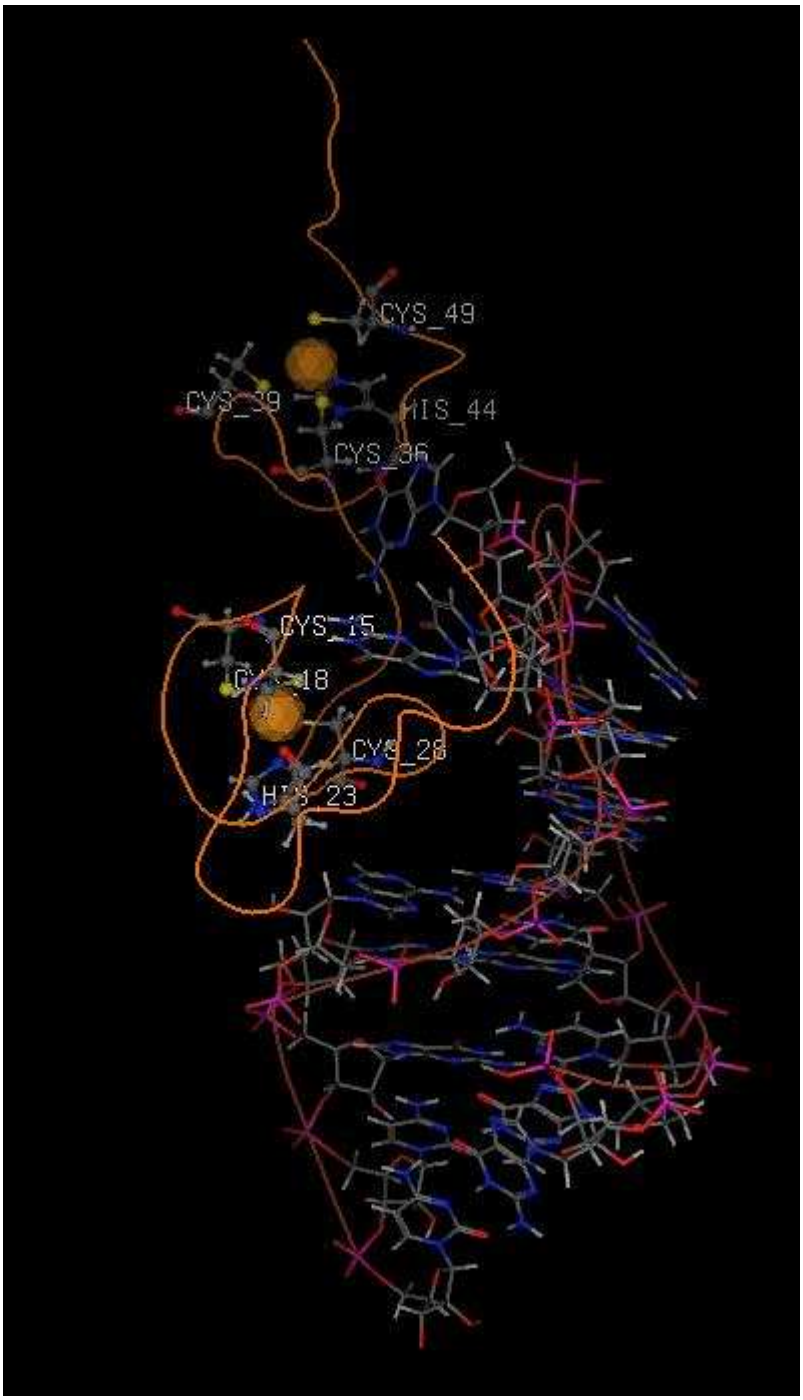


Figure 7.2 HIV nucleocapsid bound to RNA (PDB 1mfs) Lee, *et al.* (1998) *J. Mol. Biol.* Orange spheres = zinc atoms, yellow balls = cysteine residues.

Figure 7.3 shows the sequence alignment of HIV-1 and HTLV-I nucleocapsids. The two nucleocapsids share only 16% homology. HTLV-I nucleocapsid has a much longer C-terminal region. This most likely has to due with specific packaging of the RNA that forms the viral core. HIV-1 has a conical core whereas HTLV-I is spherical (Coffin *et al.*, 1997).

Figure 7.3 Sequence alignment of HTLV-I and HIV-1 nucleocapsid proteins.

Little is known about HTLV-I nucleocapsid. However, HIV-1 nucleocapsid has shown promise as a possible candidate for inhibitors (Stephen *et al.*, 2002; Turpin *et al.*, 1999). Determination of the structure of HTLV-I nucleocapsid may be crucial to finding inhibitors for alternative treatment of HTLV-I infection.

7.2 Expression of HTLV-I Gag Precursor and Nucleocapsid

HTLV-I genome K30 plasmid was generously provided by Dr. Michele Owen from the Centers for Disease Control & Prevention (CDC) and by Dr. Thomas J. Kindt from the National Institutes of Health (NIH). The HTLV-I *gag* precursor gene, in the K30 plasmid, is contained within the reading frame of base pairs 804 to 2099 (Zhao *et al.*, 1993). Primers that contain *Xho*I and *Nde*I (New England Biolabs) restriction sites were designed to express the *gag* gene and the nucleocapsid gene (base pairs 1838-2099) from the HTLV-I genome (Table 7.1) and amplification was performed by polymerase chain reaction (PCR). Since the 5' end of the nucleocapsid gene is synonymous with the 5' end of the Gag precursor polyprotein gene, the reverse Gag primer was used for the nucleocapsid expression plasmid as well as Gag expression plasmid. The amplified products were isolated from 1% agarose using the Qiagen system, digested with *Xho*I and *Nde*I and subcloned into pET-28b(+) plasmid (Novagen), which contains a His-tag region for protein purification by Ni affinity (see experimental section 9.8.2 for methods).

Table 7.1 PCR primers for construction of pNC1 and pGag1. The *Nde*I are red, the *Xho*I is blue, and the ↓ is the cleavage site.

Plasmid	Primer	Sequence
pGag1	Forward	**5'-**TTTTTTCA↓TATGGGCCAAATCTTTTCCCGTAGCG**-3'**
pGag1 pNC1	Reverse	**5'-**TTTTTTC↓TCGAGTTAAACCTCCCCCCCTATG**-3'**
pNC1	Forward	**5'-**CGCCGCCA↓TATGGTTGTCCAGCCTAAAAAA**-3'**

7.3 Results

Following PCR amplification using the primers listed in Table 7.1 agarose electrophoresis gel was carried out on the purified fragments. A band the size of NC DNA was identified (261 base pairs). The Gag precursor DNA showed more than one fragment that may be the result of non-specific binding. However, a band was identified at the expected size of the Gag amplicon (1295 base pairs), but was not purified further. The NC amplicon was inserted into pET-28b(+). Following gel extraction and purification, an agarose gel showed a band containing both the pET-28b(+) and NC insert. The ligation was thus successful and a stock solution of the NC plasmid was transformed into JM109 cells was prepared. Several LB growths were carried out and attempts made to purify NC on a nickel affinity column. However, none of the protein gels indicated the presence of NC.

7.4 Conclusion

The amplification of the Gag amplicon was accomplished. However, the band identified for the Gag amplicon was not purified and no ligation was attempted. NC was successfully inserted into pET-28b(+). Protein gels after several LB growths followed by purification attempts via nickel affinity column failed to show any indication of NC. However, the NC protein may be present below the detection limit. A more reasonable

explanation is that the NC begins to fold upon expression and forms a three-dimensional structure, in which the His-tag is buried. If the His-tag is buried within the protein, it cannot bind to the nickel column and will be lost in the flow-through. Bryan Herger attempted to express NC in the Rosetta cell line. SDS PAGE data revealed that the expected protein band was present. However, he was also unable to isolate the protein. Future work with NC expression and isolation may be more efficiently done with other expression systems such as yeast.

REFERENCES

Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000). NMR Structure of the HIV-1 Nucleocapsid Protein Bound to Stem-Loop SL2 of the ψ-RNA Packaging Signal. Implications for Genome Recognition. *J.M.B.* **301,** 491-511.

Bertola, F., Manigand, C., Picard, P., Belghazi, M., and Precigoux, G. (2000). Human T-lymphotrophic virus type I nucleocapsid protein NCp15: structural study and stability of the N-terminal zinc-finger. *Biochem. J.* **352,** 293-300.

Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997). "Retroviruses." Cold Spring Harbor Laboratory Press, Plainview.

Lee, B. M., De Guzman, R. N., Turner, B. G., Tjandra, N., and Summers, M. F. (1998). Dynamic Behavior of the HIV-1 Nucleocapsid Protein. *J. M. B.* **279,** 633-649.

Stephen, A. G., Worthy, K. M., Towler, E., Mikovits, J. A., Sei, S., Roberts, P., Yang, Q., Akee, R. K., Klausmeyer, P., McCloud, T. G., Henderson, L., Rein, A., Covell, D. G., Currens, M., Shoemaker, R. H., and Fisher, R. J. (2002). Identification of HIV-1 Nucleocapsid Protein: Nucleic Acid Antagonists with Cellular Anti-HIV Activity. *Bioche. Biophys. Res. Commun.* **296,** 1228-1237.

Turpin, J. A., Song, Y., Inman, J. K., Huang, M., Wallqvist, A., Maynard, A., Covell, D. G., Rice, W. G., and Apella, E. (1999). Synthesis and Biological Properties of Novel Pyridinioalkanoyl Thiolesters (PATES) as Anti-HIV-1 Agents that Target the Viral Nucleocapsid Protein Zinc Fingers. *J. Med. Chem.* **42,** 67-86.

Williams, M. C., Gorelick, R. J., and Musier-Forsyth, K. (2002). Specific Zinc-finger Architecture Required for HIV-1 Nucleocapsid Protein's Nucleic Acid Chaperone Function. *P.N.A.S.* **99**(13)**,** 8614-8619.

Zhao, T. M., Robinson, M. A., Sawasdikosol, S., Simpson, R. M., and Kindt, T. J. (1993). Variation in HTLV-I Sequences from Rabbit Cell Lines with Diverse *in vivo* Effects. *Virology* **195**(1)**,** 271-274.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

The primary goals of this research are to further the understanding of the virology and enzymology of human T-cell leukemia virus type I (HTLV-I) that will lead to the development of treatments for patients infected with HTLV-I. HTLV-I is an oncogenic virus of the *Retroviridae* family (Murphy & Kingsbury, 1990) and is the causative agent of adult T-cell leukemia/lymphoma (ATL) (Gitlin et al., 1993), and tropical spastic paraparesis/HTLV-I associated myelopathy (TSP/HAM) (Gessain et al., 1985). HTLV-I has been classified as a dangerous emerging pathogen by the Centers for Disease Control & Prevention (Ewald, 1996; Satcher, 1995). HTLV-I is endemic in southwestern Japan (Gallo & Thomson, 1996; Green & Chen, 1993), as well as areas in the Carribean, sub-Saharan Africa, Seychelles Islands, Melanesia, Australia, Philippines, the Americas, south India, and southwestern Asia (Gallo & Thomson, 1996). World-wide estimates are that approximately 20 million persons are infected with HTLV-I (Dekaban et al., 1992). In the United States, there is an estimated prevalence of 0.016-0.1% HTLV-I infection among blood donors (Zucker-Franklin & Pancake, 1998). Transmission is vertical via maternal breast milk, intravenous drug use with contaminated sharps, transfusion with infected blood, and sexual transmission (Gessain, 1996; Kaplan & Khabbaz, 1993; Khabbaz et al., 1990).

Currently, there is no adequate treatment for HTLV-I infected individuals. Combination therapies have been tried, with limited success. The median survival time of infected patients with acute ATL is 7-8 months (Matsushita et al., 1999). The fact that there are currently no effective treatments to control HTLV-I infection and prevent or treat HTLV-I induced ATL and TSP/HAM is obviously a significant problem.

HTLV-I protease is necessary for retroviral maturation and replication and is, therefore, an attractive target for inhibitor design. Understanding the structure and enzymology of HTLV-I protease is important for the design of inhibitors. The last ten residues of HTLV-I protease are only found in leukemia virus proteases. It was found that the last ten residues are not necessary for enzymatic activity. HTLV-I has a putative natural cleavage junction in the C-terminal region. The C-terminal cleavage appeared to be inhibited by Tyr^{114} or Glu^{117}. Studies with synthetic peptide mutants showed that the C-terminal "tail" was inhibited by Glu^{117}. HTLV-I protease mutants incorporating either a Y114V or E117V mutations were incubated at 37 °C and cleavage products confirmed by MALDI. The data confirm the Glu^{117} inhibition. In addition, a second autoprocessing site was found at Leu^{78}.

Investigation of peptide mimetic compounds incorporating the tetrahedral intermediate of aspartyl protease catalyzed cleavage are crucial for the development of lead inhibitors. Compounds containing statine (Sta), 4-amino-3-hydroxy-5-phenylpentanoic acid (AHPPA), or hydroxyethylamine (HEA) were expected to be potent protease inhibitors with $K_i < 1$ nM.

This was not the case. Our best inhibitor, the MA/CA statine-based inhibitor (Ac-QV-Sta-PV) had the inhibition constant ($K_i = 29 \pm 4$ nM). This was comparable to the CS-I-25 pepsin inhibitor with $K_i = 30 \pm 3$ nM. Percent inhibition was also similar for these

two inhibitors with 88% inhibition for Ac-QV-Sta-PV and 86% for CS-I-25. These peptide inhibitors still provide preliminary information that can be used to develop non-peptide inhibitors that have better bioavailability and pharmacokinetics. In addition, modifications of the HEA inhibitors studied here could be developed that are more efficient inhibitors. Akaji et al. synthesized an hydroxyethylamine inhibitor with a $K_i = 38 \pm 8$ nM (Akaji et al., 2003), which is in the experimental range of our statine compound.

In vitro cell culture of HTLV-I protease inhibition began to show slight inhibition after five weeks. Both the Ac-QV-Sta-PV and the Ac-QV-AHPPA-PV showed inhibition with concentrations of 5 and 10 μM. For both classes of inhibitors, the 20 μM concentrations did not appear to have any effect. However, the cell culture assays will monitored continuously to determine whether the peptides actually inhibit the virus or if results are due to cytotoxicity. In either case, the results are significant toward our understanding of antiviral therapy for of the HTLV-I.

Understanding of HTLV-I protease structure and enzymology is crucial for the design of inhibitors. The last ten residues of HTLV-I protease are only found in leukemia virus proteases. HTLV-I has a nature cleavage junction at the last ten residues in the C-terminal region. When we have the final LC-MS data on the C-terminal tail cleavage, this will confirm that the C-terminal tail is not cleaved due to the electrostatic effects of Glu^{117}. It is believed that this C-terminal tail is an artifact as a result of a point mutation. *In vitro* studies with our HTLV-I mutants could provide us with further understanding of the effects of protease enzymology with and without the last ten residues.

CHAPTER 9

EXPERIMENTAL

9.1 HTLV-I Protease C-Terminal Cleavage Junction

9.1.1 Preparation and Purification of HTLV-I Protease

Chemicals. Imidazole was purchased from Sigma, and Tris, urea, sodium acetate, and Luria-Bertani (LB broth) from Fisher. All chemicals were of reagent grade and used without further purification. Microsep concentrators were purchased from Pall Filtron Corporation. The water used in all experiments was deionized and filtered using a NANOpure Infinity base unit from Barnstead/Thermolyne Corporation.

Protease purification. The construct pPR106 containing the DNA encoding HTLV-I protease (Shuker et al., 2003)was transformed into *E. coli* BL21(DE3) (Novagen) and expressed by growth in LB (Miller) liquid media inoculated with 500 μL ampicillin (100 mg/mL). The cells were grown at 37 °C to an optical density at 600 nm (OD_{600}) of 0.6-0.8 followed by induction with isopropyl β-D-thiogalactoside (IPTG) to 0.4 mM final concentration. After incubation for 3 h, the cells were harvested by centrifugation at 10,000*g* using a DuPont Sorvall RC-5B centrifuge for 5 min at 4 °C. The pellets were

either stored at – 80 °C or resuspended in buffer A (20 mM Tris, pH 7.9, 5 mM imidazole, and 500 mM NaCl) then sonicated on ice followed by centrifugation at 10,000*g* for 10 min. The insoluble fraction was collected by centrifugation at 10,000*g* for 15 min. These steps were repeated. The pellet was resuspended in buffer B (buffer A augmented with 8 M urea) and incubated at 4 °C for 1 h. The mixture was separated by centrifugation at 14,000*g* for 45 min. The supernatant was loaded on a 4 mL His-Bind resin (Novagen) column charged with 0.5 M nickel sulfate. The resin-bound protease was washed with buffer B followed by buffer C (20 mM Tris, pH 7.9, 20 mM imidazole, 500 mM NaCl, and 8 M urea). The protease was eluted with buffer D (20 mM Tris, pH 7.9, 1 M imidazole, 500 mM NaCl, and 8 M urea). Purified protease was dialyzed in 10 mM acetic acid (pH 5.3).

9.1.2 Incubation of Peptides with HTLV-I Protease

C-terminal cleavage site peptide mutants. Two single mutant peptides pertaining to Y114V and E117V (NH_2-VLVLPEAKGP-OH and NH_2-VLYLPVAKGP-OH, respectively) and one double mutant containing both Y114V and E117V (NH_2-VLVLPVAK-OH) were purchased from SynPep. Four mg samples of each peptide were dissolved in 100 µL dimethyl sulfoxide (DMSO) then 100 µL of 10 mM NaOAc, pH 5.2 buffer (Buffer N) containing 1 mM dithiothreitol (DTT) was added. This was followed by the addition of 500 µL of 85 µM protease in Buffer N. The solution was incubated for 3 h at 37 °C and

then passed through a MicroSep concentrator membrane (Pall Gelman) with a 3000 Da cutoff at 5000*g* for 3 h to further purify the protease. The flow-through solution containing cleaved C-terminal peptides was analyzed by LC-MS.

9.2 Solid Phase Synthesis

9.2.1 Synthesis of Statine and AHPPA Inhibitors

Solid Phase Synthesis of Compounds **1**-**6** *using ArgoPore-NH_2 Resin*. AgroPore-NH_2 resin (Argonaut Technologies Inc.) with a reported substitution of 0.75 mmol/g resin was used (Wellings & Atherton, 1997). To each column, HMBA was attached to the bead as a cleavable linker. Then the Fmoc-protected amino acid was attached to the linker by forming the symmetrical anhydride. To prepare the Fmoc-protected amino acid for attachment, ten equivalents of the Fmoc-amino acid were placed in a round bottom flask in N,N-dimethylformamide (DMF). Five equivalents of N,N'-dicyclohexylcarbodiimide (DCC) were added and the mixture was stired under N_2 gas for 20 min at 0 °C. The solution was filtered and added to the HMBA-linked beads in each column. A catalytic amount of 4-dimethylaminopyridine (DMAP) was added and the reaction was allowed to continue over night on the rotisserie. The Fmoc was cleaved with 20% piperidine in DMF

(mechanism shown in Figure 5.4). Following deprotection, the next amino acid was added by base catalysis with DIPEA and the coupling agent TBTU for the ester-linked amino acids and HBTU for subsequent amino acids. Briefly, 2.5 molar equivalents of the Fmoc-amino acid and 3 molar equivalents of the coupling reagent (TBTU or HBTU) were dissolved in a minimal amount of DMF. Then this solution was added to the resin followed by the addition of 5 molar equivalents of DIPEA. The reaction was sealed and rotated for no less then 3 hours. Each coupling step was repeated two times. After coupling, the beads were rinsed with DMF followed by DCM. The Fmoc protecting group was cleaved with two successive 20 minute treatments of the resin with a 20% solution of piperidine in DMF. Efficiency of coupling was monitored via a 2,4,6-trinitrobenzencesulfonic acid (TNBS) test between each step (Hancock & Battrsby, 1976; Wellings & Atherton, 1997). Capping was performed in between each coupling with a solution of 5% acetic anhydride and 4 equivalents of DIPEA in DMF. Upon completion of synthesis, the side-chain protecting groups were cleaved with a solution of 95% TFA, 2.5% phenol, and 2.5% H_2O for 25-min. Peptides were cleaved from the beads with 50% ethylamine in acetonitrile.

Solid Phase Synthesis of Compounds **1**-**6** *using NovaSyn® TG HMBA Resin*.

NovaSyn® TG HMBA Resin (Novabiochem) with a reported substitution of 0.69 mmol/g resin was used (Sheppard, 1982). To each bead, the Fmoc-protected amino acid was attached to the linker and deprotected as described above. Following deprotection, the next amino acid was added by base catalysis. Briefly, 2.5 molar equivalents of the Fmoc-amino acid, 4 molar equivalents of the coupling reagent TBTU, and 4 molar equivalents of N-hydroxybenzotrialoze (HOBt) were dissolved in a minimal amount of

DMF. Then this solution was added to the resin followed by the addition of 5 molar equivalents of DIPEA. The reaction was sealed and rotated overnight. The remainder of the synthesis followed the same procedure as for the Argopore-NH_2 resin, except that peptides were cleaved from beads with 50% methanol in triethylamine (TEA).

Solid Phase Synthesis of Compounds **1**-**6** *using NovaSyn® TGA Pre-loaded Resin*. *NovaSyn® TGA N-α-Fmoc protected-Val(Ile) resin* (Novabiochem) with a reported substitution of 0.24 mmol/g for Val-substituted resin and 0.21 mmol/g for Ile-substituted resin was used (Sheppard, 1982; Bayer, 1991). The beads were deprotected with 20% piperidine. Following deprotection, the remainder of the synthesis followed the same procedure as for the NovaSyn® TG resin except the side-chain protecting groups were cleaved along with release of peptides from beads with a solution of 95% TFA, 5% H_2O for 30-min. Results were confirmed by LC/MS.

9.2.2 Synthesis of HEA Inhibitors

Compound **13**. A trial reaction to link TPCK to Pro was first carried out to confirm that the reaction between TPCK and Pro would proceed prior to addition of TPCK to the Pro-linked beads. For the trial reaction, to a 25 mL boiling flask, 1 equivalent of TPCK was added to 25 mL DMF along with NaI (2 eq). Proline (2.5 eq) was then added followed by $NaHCO_3$ (2.5 eq) and DIPEA (5 eq). The components did not dissolve entirely. Water was added until all the components dissolved. The reaction was allowed to stir for 48

hours. Comfirmatiion of the synthesis of compound **13** was verified by comparison of ^{1}H NMR chemical shifts for the TPCK starting material (Figure 9.1 Compound 10) and for the TPCK-Pro product (Figure 9.2 Compound 13). There was an upfield shift in the methylene protons from δ 4.25 for **10** to overlapping the β protons at δ 3.42 for **13** and the appearance of the hydroxyl on the proline at δ 10.60 for **13**.

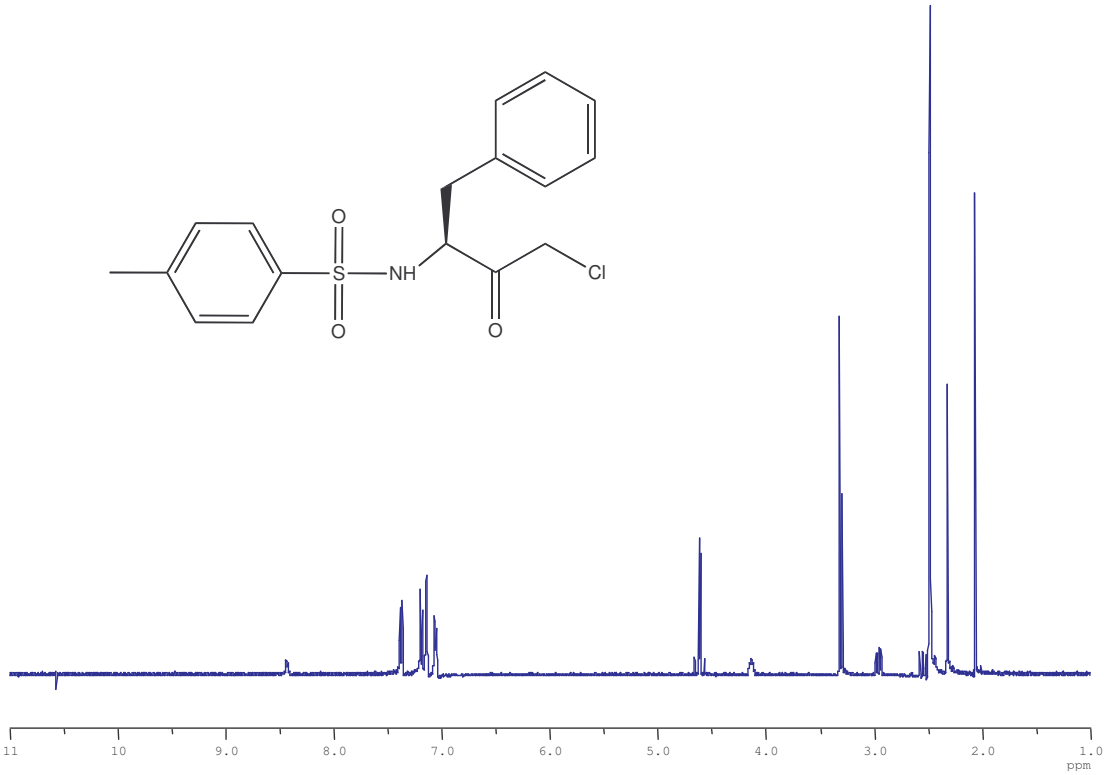


Figure 9.1 ^{1}H NMR spectrum of starting material TPCK. Chemical shift assignments: δ 7.31-7.45 (m, 9 H, Ar*H*), δ 4.52 (t, 1H, NHC*H*CH_2), δ 4.25 (dd, 2H, COCH_2Cl), δ 3.35 (acetone), δ 2.52-2.90 (dd, 2H, CH_2CH), δ 2.50 (DMSO), δ 2.35 (s, 3H, C*H*3Ar), δ 2.10 (s, 1H, N*H*).

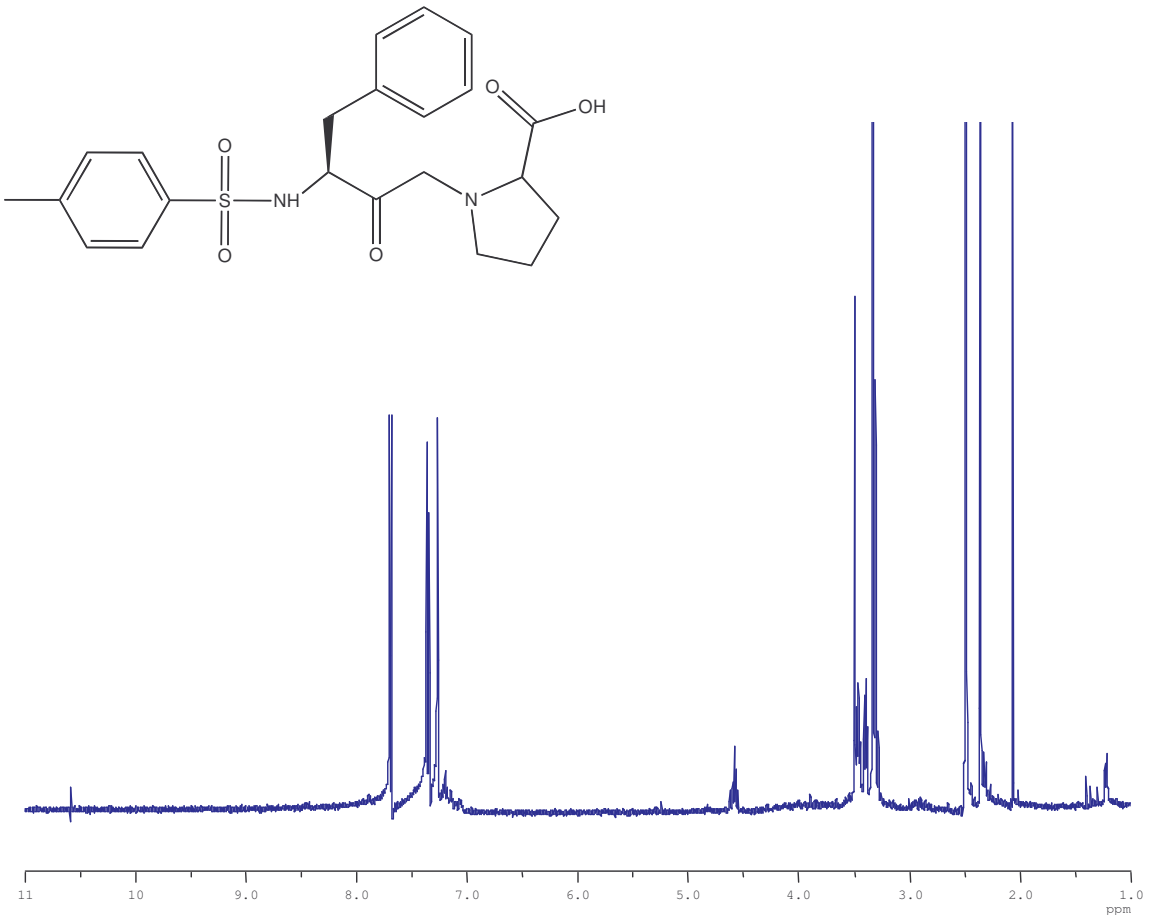


Figure 9.2 ^{1}H NMR spectrum of Tos-TPCK-Pro (**13**) product. Chemical shift assignments: δ 10.60 (s, 1H, COO*H*), δ 7.25-7.56 (m, 9H, Ar*H*), δ 4.52 (t, 1H, NHC*H*CH_2), δ 3.50 (s, 1H, NHC*H*COOH), δ 3.42 (dd, 2H, COCH_2N), δ 3.35 (Acetone), δ 2.50 (DMSO), δ 2.35 (s, 3H, C*H*3Ar), δ 2.10 (s, 1H, NH).

Solid phase synthesis of HEA inhibitors 7-8. The HEA inhibitors were synthesized as above for compounds 1-6 on NovaSyn® TGA N-α-Fmoc protected-Val(Ile) resin (Novabiochem) for to the proline residue (Pro-Val(Ile)-HMBA-❍). Next, 1 equivalent of TPCK was added along with 1 mL DMF and NaI (2 eq). Following the synthesis for Comound 13, proline (2.5 eq) was added followed by $NaHCO_3$ (2.5 eq) and DIPEA (5 eq). Water was added to dissolve the components. However, they did not dissolved entirely. The reaction was allowed to mix on the rotisserie for 48 hours. Once the reaction was completed, the TPCK was removed. This was accomplished by adding the beads to a small round bottom flask in ethanol. Ten equivalents of $NaBH_4$ and 4 equivalents of 1,2-dimethoxynaphthalene were added. The mixture was stirred overnight under N_2 gas and a 600 watt sunlamp was used as the light source for the electron transfer reaction. TPCK removal was confirmed by a TNBS test and the synthesis of the peptides were continued in the same manner as for the statine and AHPPA inhibitors. Results were confirmed by LC/MS.

9.3 FRET Inhibition Assays

Inhibition Assays. Inhibitors were assayed with varying concentrations of Dabcyl-PQVLPVMH-EDANS, a FRET substrate that absorbs at 340 nm and fluoresces at 490 nm on cleavage. HTLV-I protease (5 nM) was incubated with inhibitor (1-5 μM) for 5-20 min at 37 °C on a 96-well Costar plate in 300 μL volumes containing 10% DMSO and Buffer N (pH 5.2, 400 mM NaOAc). The substrate was added (10-20 μM) and the

fluorescence was read on a ULTRA plate reader (TECAN) with excitation 340 nm. The background correction was done using a well with that contained substrate, DMSO, and buffer. An example of the results is shown in Figure 9.3. Initial time points during which the solutions were coming to equilibrium were eliminated.

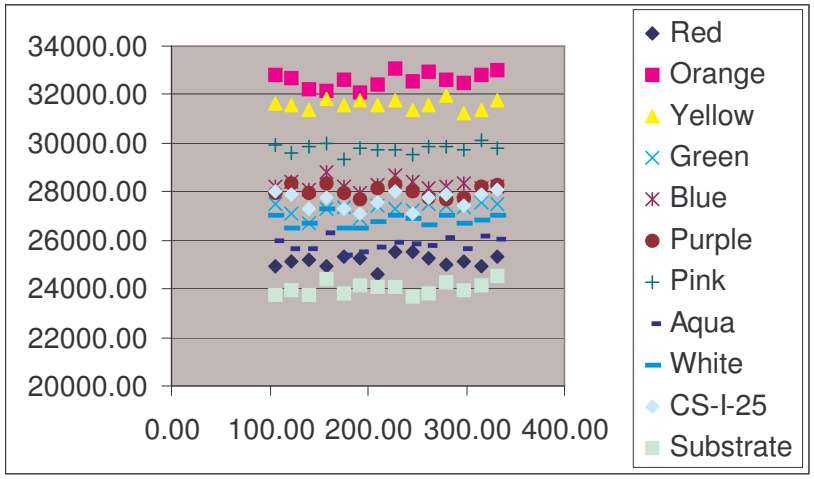


Figure 9.3 Data showing increasing fluorescence of inhibitors and substrate. Inhibitors are color-coded in the following manener: Red (Ac-QV-Sta-PV), Orange (Ac-QV-AHPPA-PV), Yellow (Ac-SI-Sta-PV), Green (Ac-SI-AHPPA-PV), Blue (Ac-VI-Sta-PI), Purple (Ac-VI-AHPPA-PI), Pink (Ac-QV-F-HEA-PV), Aqua (Ac-SI-F-HEA-PV), and White (Ac-VI-F-HEA-PI).

The velocity was found by taking the slope of each line for each inhibitor. Percent inhibition was calculated by dividing the intensity for the substrate in the presence of inhibitor by the intensity for the substrate alone using the following equation: $(1-(F_i/F_s)*100)$, where F_i is the fluorescence intensity of the substrate in the presence of inhibitor and F_s is the fluorescence intensity of the substrate alone.

The inhibition constants were determined by a method developed by Dixon (Segal, 1975) where substrate and inhibitor concentrations are varied and plotted as 1/V vs. [Inhibitor]. An example of a data set is shown below in Figure 9.4.

		velocity				1/V			
	[Substrate]	10 uM	15 uM	17 uM	20 uM	10 uM	15 uM	17 uM	20 uM
	0	0.895	1.030	1.090	1.150	1.12	0.97	0.92	0.87
	1	0.111	0.715	0.648	0.256		1.39	1.54	3.91
[Inhibitor]	1.33	0.344	0.323	0.350	0.112	2.91	3.10	2.86	8.93
uM	3	0.166			0.452	6.02			
	5	0.086	0.370	0.478		11.63			20.33
	2		0.180	0.826			5.56		

Figure 9.4 Data set for Ac-QV-Sta-PV kinetics.

An example of a resultant Dixon plot is shown in Figure 9.5.

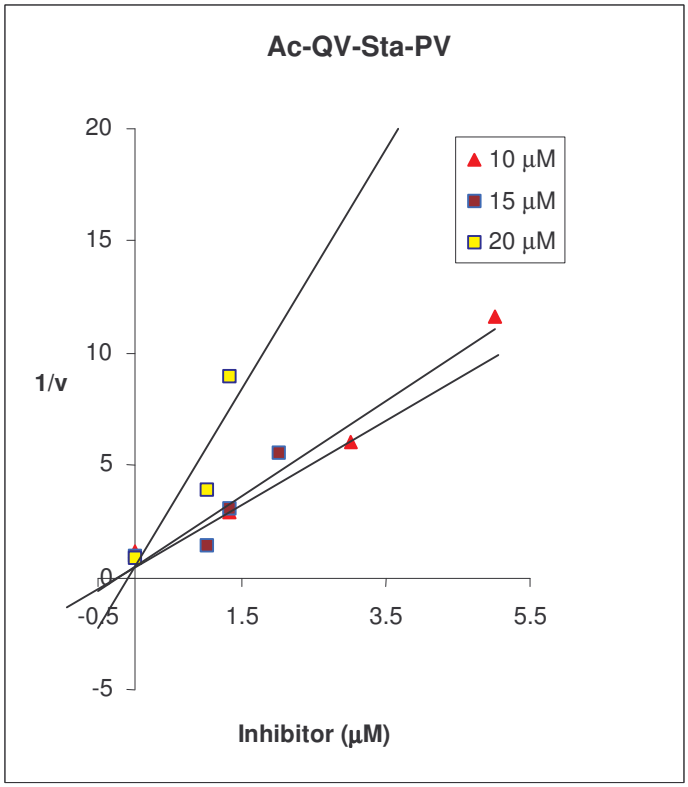


Figure 9.5 Dixon plot for Ac-QV-Sta-PV inhibition.

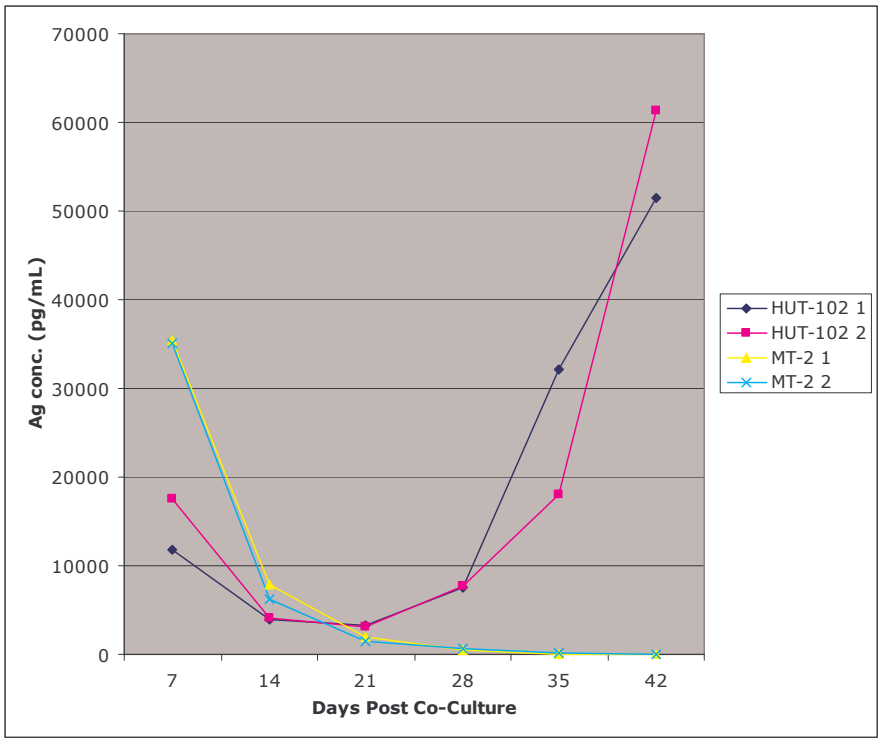
9.4 HTLV-I Infection

Cell Culture. Human T-cell lines from Hut-78 cell were cultured in a 5% CO_2 incubator at 37 °C using RPMI-1640 media supplemented with 10% fetal calf serum, 100 U/mL penicillin, 100 µg/mL streptomycin, and 2 mM L-glutamine (Gibco Laboratories, Grand Island, NY). Cell cultures were split 1:10 twice weekly. This protocol was done to insure

that proper handling and maintenance of cell lines could be accomplished before working with HTLV-infected cell lines.

Co-Infection of PBMCs with Hut-102 and MT-2 Cell Lines infected with HTLV-I. Three days prior to setting the infection, two vials of normal, uninfected donor PBMCs were thawed. Cells were counted and placed in a in a T150 flask at 10^6 cells/ml in RPMI-10% FCS, 10% IL-2. Phytohemaglutinin (PHA) to a concentration of 0.1% PHA (1µl PHA / ml of medium) was added to stimulate growth of PBMCs. The flask was incubated for 3 days to activate the PBMCs. The cell number was doubled during the three days to an approximate concentration of $2x10^6$/ml. Approximately 3 mL of PHA-blasted PBMCs were removed from the T150 flasks and centrifuged at 1200 RPM for 10 minutes. The supernatants were removed and resuspended in fresh medium so that the cells were again at 10^6 cells/ml in RPMI-10% FCS, 10% IL-2. HTLV-I-infected cells (HUT-102 and MT-2) were cultured in RPMI-10% FCS. The actively growing HUT-102 and MT-2 cells were counted using trypan blue in a hemacytomoter to determine cell viability. Cells were centrifuged and resuspended at 10^6 cells/ml in RPMI-10% FCS, 10% IL-2. The HUT-102 and MT-2 cells were irradiated with 6,000 rads using a gamma cell irradiator. To a 24-well plate, PHA-blasted normal PBMC cells were added (10^6 cells/well in 1 mL). HUT-102 and MT-2 (HTLV-I-infected cells) were added to the normal donor PHA-blasted PBMCs in a 1:1 ratio for a total volume of 2 mL/well. In addition, one well was set containing only irradiated HUT-102 cells to monitor cell death. The cultures were maintained in RPMI-10% FCS, 10% IL-2 and supernatants were collected every 7^{th} day to monitor infection by p19 antigen ELISA. The results of the first trial infection are shown in Table 9.1.

Table 9.1 HTLV-I infection of PBMCs



The results in Table 9.1 indicate that the HUT-102 infections were successful as is seen by the re-emergence in p19 antigen absorbance after 21 days. The MT-2s do not show the increase in p19 antigen and thus the infection of MT-2 was not successful. The declining curve of the MT-2s show that the irradiated MT-2 cells die off without infection of PBMCs, which would show an increase in the p19 antigen as is shown for the HUT-102 curves. However, the MT-2 data do show results that should be seen if the protease inhibitors are effective.

9.5 Inhibition of HTLV-I Protease in Cell Culture

Inibition of HTLV-I in infected culture. The HUT-102 PBMC co-culture experiment was set up as previously described with the addition of 5, 10, or 20 μM concentrations of Ac-QV-Sta-PV, CS-I-25, or with Leu-Leu-Leu tripeptide as a control. Inhibitors were diluted into RPMI from 10 mM stock solutions in DMSO. Supernatants were collected to monitor status of infection and fresh media replaced with inhibitor. Inhibitor concentrations were maintained throughout the culture and incubated at 37 °C.

9.6 HTLV-I Protease Models

Generation of Models. The programs MODELLER (Sali & Blundell, 1993; Marti-Renom et al., 2000; Fiser et al., 2000)] and Molecular Operating Environment (MOE) (Chemical Computing Group, Inc.) were used to generate structural models of HTLV-I protease with and without the ten C-terminal residues, using the sequence alignment generated by MODELLER and Protein Data Bank (PDB) structure files 1BAI (Wu et al., 1998) and 1D4L (Tyndall et al., 2000).

9.7 Fluorescence Quenching of HTLV-I ProteaseTryptophan Residues

Fluorescence quenching of HTLV-I tryptophans. Acrylamide quenching measurements at constant pH were carried out by monitoring the effect of increasing amounts of acrylamide (0-0.4 M) on solutions of HTLV-I protease in sodium acetate buffer (10 mM, pH 5.3). The KI quenching was similarly performed except that constant ionic strength was maintained using NaCl. A small amount of sodium thiosulfate (0.1 mM) was added to the KI solutions to prevent I_3^- formation, which absorbs within the wavelength region of tryptophan fluorescence (Eftink, 1991). However, this did not seem to have any effect on I_3^- formation. Fluorescence measurements were carried out on a Shimadzu RF 5301 PC spectrophotometer. Excitation was at 295 nm and fluorescence emission was scanned from 300 to 500 nm. Slit widths of 5.0 nm were set for both excitation and emission in all experiments. The data were corrected for background buffer fluorescence in the presence of 0 and highest quencher concentration. Dilution corrections were used to account for the addition of quencher. Finally, absorbance corrections were also made by taking the absorbance at 295 nm for HTLV-I protease with 0 and maximum amount of quencher. By using the following equation: $I_{corr} = 10^{\frac{Abs295nm}{2}} I_{measured.}$ the absorbance standard curve could be applied to a corrected fluorescence intensity.

9.8 Expression and Purification of HTLV-I Nucleocaspid

9.8.1 Materials

Plasmids. The following reagent was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HTLV-I K30 DNA from Dr. Thomas Kindt (Zhao et al., 1995). The pET-28b(+) vector was purchased from Novagen.

Enzymes. Restriction endonucleases, Vent DNA polymerase, T4 DNA ligase, and calf intestinal alkaline phosphatase (CIP) were purchased from NEW England BioLabs.

Oligos. Oligodeoxyribonucleotides were obtained from Stratagene.

Primers. All primers were purchased from Integrated DNA Technologies.

9.8.2 Methods

Two PCR reactions were set up to amplify the NC and Gag precursor DNA sequences from the HTLV-I K30 genome. In one microfuge tube, 10 μL 10X PCR buffer, 100 mM $MgSO_4$, 10 μL HTLV-I K30 template, 1 μL forward Gag primer (Table 7.1), 1 μL reverse Gag primer, 1μL dNTP, 0.5 μL DNA vent, and 74.5 μL deionized water were added. In a separate microfuge tube, 10 μL 10X PCR buffer, 100 mM $MgSO_4$, 10 μL HTLV-I K30 template, 1 μL forward NC primer, 1 μL reverse Gag primer, 1μL dNTP, 0.5 μL DNA vent, and 74.5 μL deionized water were added. The amplified DNA was purified using the Qiagen prep kit. An agarose gel was run to check for the 53 kDa Gag precursor and 9.5 kDa NC fragments. A 39 μL aliquot of the NC DNA was then incubated overnight in a solution of 5 μL 10X Buffer 2, 5 μL 10X BSA, and 1 μL of the *Xho*I restriction endonuclease. Another tube containing 39 μL of pET-28b(+) vector, 5 μL 10X Buffer 2, 5 μL 10X BSA, and 1 μL of the *Xho*I endonuclease were simultaneously incubated overnight. Both solutions were heat shocked at 70 ^{o}C for 20 min. followed by centrifugation. 1 μL of CIP was added to the pET-28b(+) solution to prevent self-ligation. Both solutions were incubated at 37 ^{o}C for 1 hr. Following incubation, the DNA was isolated on an agarose gel, extracted, and purified by the Qiagen prep method. Another agarose gel was run with the purified solutions to ensure that NC and pET-28b(+) DNA were present. Three microfuge tubes were labeled L1(ligation 1), L2 (ligation 2) and, LC (control). Reagents were added to the tubes according to Table 9.2.

Table 9.2 Reagent for pNC1 ligation. Two reactions (L1, L2) and a control (LC) were carried out.

Reagent	L1 (μL)	L2 (μL)	LC (μL)
pET-28b(+)	10	5	10
NC DNA insert	7	12	
dH_2O	-	-	7
10X T4 buffer	2	2	2
dNTP	0.5	0.5	0.5
T4 ligase	0.5	0.5	0.5
Total volume	20	20	20

All microfuge tubes containing the above solutions were incubated overnight at 4 ^{o}C. The pET-28b(+) vector containing the nucleocapsid nucleotide sequence was named pNC1. An agarose gel was run to verify ligation. The pNC1 vector was transformed into JM109 cells. The JM109 cells were thawed on ice for 5 min. Three microfuge tubes were labeled L1, L2, and LC. Four μL of pNC1 was added to the JM109 cells while set on ice. The cells were heat shocked at 42 ^{o}C for 40 s. To each of the three microfuge tubes, 800 μL SOC was added followed by 200 μL JM109/pNC1solution. The solutions were incubated for 1 hr at 37 ^{o}C. Simultaneously, 3 Luria-Bertani (LB) agar plates containing kanimycin (Kan) (50 μg/mL) were incubated for 1 hr at 37 ^{o}C. Each culture plate was plated with 200 μL of cell/vector media and incubated overnight at 37 ^{o}C.

REFERENCES

Bayer, E. 1991 *Angew. Chem.*, **103,** 117.

Eftink, M. R. 1991. Fluorescence quenching reactions: probing biological macromolecular structures. In: *Biophysical and Biolochemical Aspects of Fluorescence Spectroscopy* (Ed. by Dewey, T. G.), pp. 1-41. New York: Plenum Press.

Fiser, A., Do, R. K. & Sali, A. 2000. Modeling of loops in protein structures. *Protein Sci*, **9,** 1753-73.

Hancock, W. S. & Battrsby, J. E. 1976. Anal. Biochem. *71*, **261**.

Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, **29,** 291-325.

Sali, A. & Blundell, T. L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, **234,** 779-815.

Segel, I. H. 1975. *Enzyme kinetics: Behavior and analysis of rapid equilibrium and steady-state enzyme systems*. New York: John Wiley and Sons.

Sheppard, R. C. 1982. *Int. J. Pept. Protein. Res.*, **20,** 451.

Shuker, S. B., Mariani, V. M., Herger, B. E. & Dennison, K. J. 2003. Understanding HTLV-I Protease. *Chem. Biol.*, **5,** 373-380.

Tyndall, J. D., Reid, R. C., Tyssen, D. P., Jardine, D. K., Todd, B., Passmore, M., March, D. R., Pattenden, L. K., Bergman, D. A., Alewood, D., Hu, S. H., Alewood, P. F., Birch, C. J., Martin, J. L. & Fairlie, D. P. 2000. Synthesis, stability, antiviral activity, and protease-bound structures of substrate-mimicking constrained macrocyclic inhibitors of HIV-1 protease. *J Med Chem*, **43,** 3495-504.

Wellings, D. A. & Atherton, E. 1997. Standard Fmoc Protocols. *Meth. in Enzym.*, **289,** 44-66.

Wu, J., Adomat, J. M., Ridky, T. W., Louis, J. M., Leis, J., Harrison, R. W. & Weber, I. T. 1998. Structural basis for specificity of retroviral proteases. *Biochemistry*, **37,** 4518-26.

Zhao, T. M., Robinson, M. A., Bowers, F. S. & Kindt, T. J. 1995. Characterization of an infectious molecular clone of human T-cell leukemia virus type I. *J. Virol.*, **69,** 2024-2030.